www.jusletter-it.eu

David Rosenthal / Livio Veraldi

Training AI language models with third-party content and data from a legal perspective

Large language models require substantial language content for their training. Much of this content comes from the internet. However, this content may contain personal data and intellectual property of third parties. Under what conditions is its use for training purposes permissible? To answer this question, we first analyse the purpose and functioning of the training of large language models. We then consider the relevant training content sources from a legal perspective. Finally, we explain how these sources are handled under Swiss copyright, data protection, unfair competition and contract law. We also cover «crawler bans» and provider liability.

Category of articles: Articles
Field of law: IT-Law, IP Law, Competition law, Data Protection

Citation: David Rosenthal / Livio Veraldi, Training AI language models with third-party content and data from a legal perspective, in: Jusletter IT 25 March 2025

ISSN 1664-848X, jusletter-it.weblaw.ch, Weblaw AG, info@weblaw.ch, T +41 31 380 57 77 EDITIONS WEBLAW

Contents

I. Purpose of the training

[1] The training of a large language model consists of a system analysing existing texts word by word (or, in fact, token by token) to derive information on how language can be generated artificially. It is not about the possibility of preserving specific texts in order to be able to «remember» them later or even to reproduce them as such, i.e., in more or less the same form (so-called memorisation, which in this case would be verbatim).

[2] However, memorisation cannot be avoided if the same text is seen frequently enough during training, because the text then forms part of the language knowledge as such (and not just the linguistic information gained from it). This language knowledge is used in the application to continue the user's input with matching text. This is then the output. If the model has seen a sentence often enough during training («The apple does not fall far from the tree»), it will apply this and if given the input («The apple does not fall far . . . »), will continue with the phrase it has learned because this is statistically the most likely continuation for the model. For example, a large language model can memorise characteristic slogans and reproduce them easily if triggered accordingly.

[3] The same applies to factual knowledge: The main goal of training a large language model is not that the model memorises the factual knowledge from a single content in exactly the same form. Nevertheless, if the model sees certain factual information frequently enough during training (statistic relevance), it will become part of the model's factual knowledge. This is intended. This knowledge can be general knowledge (such as how our world works), but it can also include specific knowledge, as can be found in the specialised literature as a whole. It is also the reason why large language models, for example, know personal data of some famous people such as the birthday of Donald Trump very well, but they can only guess the exact birthday of another public figure who has barely made it public and therefore it could not be included in the training – which they do because it is part of their job to produce linguistically correct sentences and therefore also complete birthday details.

[4] In practice, *verbatim memorisation* in large language models is counteracted by trying to use each source document only once for training. To do this, so-called de-duplication is carried out (sometimes even line by line). This is also a good practice, as the language model does not improve if it sees the same content multiple times – it merely becomes more fixated on it, which ultimately also harms its performance. However, the aim is to obtain as diverse and comprehensive a picture of language as possible, i.e., to depict rules, meanings and other associations of language elements in the language model that are universal because they occur in different content. Of course, certain content or text fragments that are popular can occur in different documents. If a certain slogan is frequently used in the language, it also makes sense for the large language model to use it in the corresponding context, because it wants to build its outputs according to the same principles and rules as the language in the training content. This is the purpose of a large language model.[1]

[1] For the sake of completeness, it should be noted that the concept of memorisation is merely a simplified representation of the processes in a large language model in order to discuss and classify certain legal and factual problems. This is because there is no universally valid point at which content is considered memorised. When we say, for example, that a language model «knows» that Harry Potter is a wizard, it is because «Harry Potter» and «wizard» are comparatively close in one of the countless dimensions in which a language model imagines the meaning and relationship of words, and therefore the probability that Harry Potter is just such a wizard is very high. But it remains only a probability. The relationship or proximity to other properties or other elements of knowledge will be less strong, but it exists and can be calculated mathematically. In practice, content is considered memorised if the model repeatedly delivers it in exactly the same way for comparable prompts and different parameters (such as «temperature») instead of varying its answers in this respect because other content is comparably probable (so-called hallucinations). Another equally relevant question is how likely it is that a corresponding prompt will be entered at all and who is responsible for this. For detailed information on this in the context of data protection: David Rosenthal, Part 19: Language models with and without personal data, available at: https://www.vischer.com/en/knowledge/blog/part-19-language-models-with-and-without-personal-data/; see also: David Rosenthal, Part 21: The right of access to large language models, available at: https://www.vischer.com/en/knowledge/blog/part-21-the-right-of-access-to-large-language-models/.

II. Technical background

[5] When training large language models, the corresponding training content must first be cleaned to make it suitable for machine learning. This requires it to be duplicated; partially redacted to comply with data protection and broken up into small pieces (so-called tokens) that can be used by the machine. This means that the relevant training content is copied and processed in various ways. In particular, short-term copies of the content are made in the working memory of the systems used and also on the other storage systems of the person training the model. However, the language model itself does not store the training content as such, but rather the «average» of the knowledge gained from it. The model itself does not create a copy of the individual training content, but basically «only» statistics relating to it.

[6] We have described elsewhere in much more detail (for non-specialists) how large language models work in detail and explained in broad terms why they are able to do what they do.[2] However, the training process of a large language model can be described in simplified terms as follows (here without differentiating between the individual components of the language model):

1. To a certain extent, the model starts «dumb» – its billions of parameters are set at random. If an input is given to the model (which will later be a prompt), a random, meaningless string comes out.
2. The model is now presented with numerous texts for basic training. This is done piece by piece. Each piece is referred to as a token. In OpenAI's popular «gpt-4o» model, for example, the English word «infringement» consists of the four tokens «in» (no. 258), «fr» (no. 1739), «ing» (no. 289) and «ement» (no. 962).[3] The language models therefore work with tokens, not with letters or words. However, nothing is lost in the process, i.e. the conversion from text to tokens and back to text is possible without loss.[4]
3. The language model is now fed with the beginning of a text (as a prompt, so to speak) and it must then calculate how the text continues. In a figurative and highly simplified sense, this can be imagined as a spreadsheet with billions of cells. Each cell passes on the values it receives to other cells based on simple formulas and pre-programmed parameters. In the first column, the input is entered (the tokens of the input). In the last column of the spreadsheet, there is a value for each token. The token with the highest value is used.
4. During training, it is now checked whether the token with the highest value is also the one that would come next in the text fed in. If this is not the case, the parameters of the preceding cells are calculated backwards so that the «correct» token would have the highest value on the last column. Based on these calculations, these parameters are then adjusted a little like screws so that it works better the next time.
5. This is repeated countless times, but each time with a new text. Techniques against memorisation are also used, in which the model only has to predict some of the words, for example,

2 More on how a large language model works technically: David Rosenthal, Part 17: What is inside an AI model and how it works, available at: https://www.vischer.com/en/knowledge/blog/part-17-what-is-inside-an-ai-model-and-how-it-works/.

3 Try it out for yourself: https://tiktokenizer.vercel.app/.

4 However, this does not mean that large language models store all training texts. They do not. Tokens are simply a more efficient way of representing text for the purposes of training. Different providers sometimes use distinct types of tokenisation.

and not every single token is used as a training target. However, the more texts that are processed, the better the model is able to supplement the texts. It is also the reason why there are only now such powerful large language models: In the past, there simply was not the computing capacity for them.

6. However, the model is not yet ready for use. In order for it to be used for a chatbot, for example, or for other tasks, it must be trimmed for such applications. This is also done through training. This training consists of providing examples, such as questions and suitable answers. However, it only requires a fraction of the volume of training content that was used for the basic structure of the language model (also known as «pre-training»). When companies like OpenAI are interested in leveraging users' use of their chat applications to train their models, it is typically this type of training (rather than basic model knowledge building) that they are interested in. During this training, the large language models turn from text completion automata into useful tools.
7. Finally, a prompt can be used to control even more precisely what a language model does. However, large language models such as «ChatGPT» do not work only with the prompt that the user enters. They also have so-called system prompts, i.e. instructions of different lengths that are given to the model together with the user's input when it is supposed to create a response.[5]

[7] It is important for this discussion that the language model does not store the training content as such. The result of the training is an analysis of existing texts in order to derive superordinate information on how language can be generated artificially. This information is stored in the language model – viewed on a meta-level – in the form of associations between language elements and patterns in billions of parameters and is not easily readable by humans; they only see the parameters. Special techniques are needed to identify the associations and thus make visible which terms are close to each other, e.g., in a certain respect or not. Each new text contributes to the adjustment of the parameters and thus to language formation in the model – at least to a small extent – by mapping the linguistic information it contains. The linguistic information is basically like individual grains of sand on a sandy beach; their mass makes up the beach. The aim of the training is for the model to become better and better at generating new texts – not reproducing training content. Finally, the training data sets are regularly stored for later retraining. Certain content is also separated out for testing, i.e. used in a downstream, similar process.

III. Sources for training

[8] There are numerous sources for training large language models today:

- **Publicly accessible websites:** Texts from publicly accessible websites and other internet resources are read and processed. This process is called web scraping. A program calls up the website in the same way as any other user, reads the text and code of the pages and extracts the actual content from it, which it stores in a database for further processing. This is done either directly by an organisation using search robots (so-called crawlers) itself, or

5 The providers normally keep the system prompts secret, but it is sometimes possible to elicit them from the chat services, e.g., here: https://patmcguinness.substack.com/p/gpt-4-system-prompt-revealed.

by using already prepared data sets from previous crawler searches. For example, the US foundation Common Crawl was founded in 2007 and has since recorded over 250 billion internet pages and made them available in processed form, including for the training of large language models. Every month, 3–5 billion websites are added. There are also other such data sets (e.g. on the well-known machine learning platform Hugging Face). What these data sets have in common is that they consist of freely accessible content. However, freely accessible does not mean that there are no third-party rights to this content. The texts in an online magazine may also be freely accessible, but they are still protected by copyright. They may also contain personal data.

A special category of freely accessible content is **stolen content** (e.g. from ransomware attacks) that is published on more or less easily accessible platforms. A well-known example is the «books3» data set from the machine learning data set collection «The Pile». It is a data set that is said to contain 196'640 books by authors such as Stephen King, Margaret Atwood and Zadie Smith and was created in 2020, but is considered a pirated text and was therefore removed from «The Pile» following a complaint in August 2023 (although it is still available on the internet).

- **Open Access platforms:** There are various platforms that can be used to make articles, books and other content accessible to the general public and available for use beyond mere consumption, using various open access licences (such as from «CreativeCommons»). One example of scientific publications (without peer review) is «arXiv», with around 2.4 million articles. Other content also includes software code that is offered under open-source licences. A well-known example is the publicly accessible offerings on «Github». On other platforms, content is offered that is considered public domain, i.e. the copyright has been relinquished or the protection has expired. One example is «Project Gutenberg». Data sets with such content are also available on Hugging Face. Books and their content that are accessible in libraries also belong in this category, at least if they have been purchased by the libraries (but not online services that can be used with a library subscription).
- **User-generated content platforms:** These include «free» platforms such as Wikipedia, but also commercial offerings such as YouTube or Reddit. Multimedia content can also be used to train language models by transcribing spoken language (e.g. in videos with instructions) as a first step. There are already providers who offer corresponding data sets (e.g. YouTubeCommons). Even if the content on these platforms originates from their users, it is subject to licence conditions that are either formulated on the platform or defined by the users themselves. Sometimes rights are also assigned to the operators of the platforms, who can then use them themselves. The content on these platforms can also affect third-party rights, e.g. if a user uploads unauthorised third-party content.
- **Contractually accessible content:** This refers to content that is not freely accessible, i.e. without registration or other agreement. This includes, for example, content from mass media for which a paid subscription or even just free registration is required, or content from online platforms and archives that are only accessible to their registered users. This can be public or non-public content. Content is «public» if it is accessible to an indefinite number of people, even if there is a payment barrier, for example. The decisive factor is the control of access, which is only granted to those who contractually undertake to comply with certain conditions.

- **Public libraries:** If a library is freely accessible to the public, i.e. without the conclusion of a contract, it does not fall under the above category. Physical libraries differ from a website, an online medium or the open access platforms mentioned in that they do not provide online copies of works but physical copies of works (e.g. printed books and journals). This is relevant from a copyright perspective because the copyright holder of a work who has consented to the sale of a copy of that work cannot prevent that copy from being resold or otherwise distributed or even made available in a library for borrowing (principle of exhaustion). However, exhaustion is limited to the right of distribution. This means that further use (e.g. scanning) of content from public libraries may be relevant under copyright law. In contrast, when a work is published on an open access platform, the copyright holder has usually consented to further use. In this case, the work itself is typically provided with a free licence, i.e. it can be used independently of the open access platform in accordance with the licence terms (or not).
- **Synthetic content:** Synthetic content is a special category of the aforementioned sources for training because it was not created by humans and is therefore not subject to copyright protection in many jurisdictions including Switzerland.[6] This refers, for example, to the outputs generated by an existing AI model such as «gpt-4o». Such outputs are sometimes used to train language models too, although this does pose special problems in terms of data quality. The use of synthetic data can also avoid data protection issues, for example when the synthetic data is used to replace «real» personal data with data randomly «invented» by the computer, which looks the same or is structured in the same way and appears «real» but does not refer to an actual living person. This has the advantage that the language model learns what this data looks like and how it is used in language, but there is no real person involved. For this reason, synthetic data is also used for testing purposes.
- **Combined content:** There is a growing number of curated data sets that can be used directly for training large language models and which already combine several of the sources mentioned above. One example is «OpenHermes, a data set combined from various sources for fine-tuning, consisting of (mainly synthetically generated) question-answer sequences.

IV. Training: Areas of law to be considered

[9] In particular, the following areas of law are affected by the training of a large language model:[7]

- **Copyright law:** If content is protected by the Copyright Act (**CopA**), it may only be used if the rights holder has either given their consent or if it is permitted by an exception to copyright (e.g. quotations). However, this raises the question of whether a specific piece of content is protected by copyright at all, and whether the use of such content for the training of a language model is a legally relevant use. In Switzerland, this only applies to intellectual creations with an individual character (whereby an individual character is not required for photographs), provided that the copyright protection period has not expired (generally

6 However, they may be relevant under copyright law because they fall within the scope of protection of existing works and may therefore constitute a copyright infringement.

7 See, to that effect: Matthias Städeli/Lisa Mary, Künstliche Intelligenz und Urheberrecht, in: SJZ 120/2024, p. 244 et seqq., p. 244.

70 years after the death of the author) or does not exist by law (e.g. acts, ordinances, decisions, minutes and reports issued by authorities and public administrations).

- **Data protection law:** Insofar as content relates to an identified or identifiable person, it is personal data. According to the Data Protection Act (**DPA**) and the EU General Data Protection Regulation (**GDPR**), the processing of such data first requires compliance with a number of principles (transparency, good faith, purpose limitation, proportionality including data minimization and storage limitation, accuracy with regard to the purpose, data security). According to the GDPR, and in Switzerland for federal bodies also according to the DPA, a sufficient legal basis is also required (e.g. consent, legitimate interest, legal provision). Furthermore, a justification is required in the private sector if a data subject objects to the processing of their data (i.e. does not want their data to be processed and specifically informs a data processor of their objection). In the DPA, this is also the case if particularly sensitive personal data is disclosed and under the GDPR if particularly sensitive personal data is processed. In the DPA, private organizations can justify the violation of the above-mentioned principles by, inter alia, proving an overriding interest.

  At this point, Art. 28 of the Swiss Civil Code (**SCC**) should also be mentioned, which can also be discussed in a broader sense because the provision also protects the personality – incidentally, unlike the DPA, this protection extends to the personality of legal persons. Insofar as the latter are affected by the training of a model, this provision could be relevant. However, no such cases are known to date.

  We do not discuss the GDPR in detail in this article. In principle, it does not apply to the training of language models in Switzerland by bodies based in Switzerland because neither of the two requirements of Art. 3 para. 2 GDPR are met.

- **Unfair competition law:** Depending on the legal system, unfair competition law complements copyright law by protecting content from a different perspective (including content that is not protected by copyright). It must therefore be examined separately. In Switzerland, the Unfair Competition Act (**UCA**) prohibits the exploitation of entrusted or unauthorised work products without the permission of the person to whom the work product belongs. Furthermore, the UCA prohibits the adoption and exploitation of another person's work product that is ready for the market by means of technical reproduction processes without any reasonable effort of their own.

- **Protection of secrets:** In some legal systems, the violation of confidentiality obligations or the use of content that someone has obtained as a result of a violation of a confidentiality obligation is punishable. In Switzerland, this is done under the UCA and the Swiss Criminal Code (**SPC**). Confidentiality obligations can arise from various circumstances, such as a contractual relationship or legal regulations such as data protection law or provisions on official secrecy.

- **Criminal law:** In the case of certain harmful content, the law prohibits dealing with it or even possessing it in whole or in part, for example in the case of child pornography or public incitement to hatred, to crime or to violence.

- **Special legislation on protection:** In individual jurisdictions, special legislation protects certain content against use without permission. In the EU, for example, databases for which substantial investments have been made are protected for 15 years against the use of all or a substantial part of their content.

- **EU AI Act:** With the regulation on artificial intelligence (**AI Act**), the EU has introduced a regulation geared towards product safety in the field of AI. It came into force on 1 August 2024 and is being implemented gradually over a period of 36 months. General-purpose AI models, including large language models, are also regulated. However, the AI Act does not prescribe the content with which such models are to be trained. Instead, it defines various documentation obligations in Art. 53(1) AI Act, including in relation to the training and testing of a model. Users of these models are to be provided with certain minimum information. However, Art. 53(1) AI Act also contains a rule that relates to copyright and could potentially be relevant in the present case. For general-purpose AI models, the provision stipulates, among other things, that a «policy to comply with Union law on copyright and related rights» must be put in place and that a «sufficiently detailed summary about the content used for training of the general-purpose AI model» must be drawn up and made publicly available (see also recital 105). The regulation also mentions a limitation provision of EU copyright law, the so-called text and data mining exception (TDM), which allows the use of content for the training of language models, whereby the rights holder has an opt-out right with regard to applications outside of scientific research.[8] The AI Act requires that it must be checked in each case whether this opt-out right has been exercised. However, the question arises as to whether the requirements of the AI Act also apply if EU copyright law itself does not «want» to be applied to a certain situation because, according to the *lex loci protectionis* (law of the country for which protection is claimed), training of a language model takes place exclusively in Switzerland or, for example, in the USA. As part of the aforementioned strategy, it must then only be ensured – as the bare minimum – that the model is firstly not trained in the EU and secondly that EU copyright is not infringed by an output of the language model if it is subsequently distributed in the EU.[9] We believe that this restrictive approach is the right one, resulting from the wording on the one hand, and on the other hand from the fact that the AI Act was not intended to change the existing copyright law, i.e. not extend its applicability.[10] This is somewhat at odds with the recital stating that the EU regulation is intended to create a «level playing field» for all providers, which could be understood to mean that the same copyright rules should apply to all organisations wishing to offer models in the EU, regardless of where the training takes place.[11] However, such an idea appearing only in a recital and arising from the ignorance or misunderstanding of those who drafted the provision, is not capable of overriding the actual regulation of the AI Act and copyright law: The AI Act only requires that a «policy» exists on how to comply with EU copyright law, and to comply with EU copyright law is only but still required if EU copyright law «wants» to be applied itself – which, as shown, is usually not the case as far as training a large language model abroad is concerned. This interpretation can also resolve the contradiction with the recital. Incidentally, the same

8 We go into this topic in more detail in the chapter V on copyright law.

9 See: excursus in chapter V.F.

10 David Rosenthal, Der EU AI Act – Verordnung über künstliche Intelligenz, in: Jusletter of August 5, 2024, SN 65, with further references.

11 Recital 106 AI Act: «[...] Any provider placing a general-purpose AI model on the Union market should comply with this obligation, regardless of the jurisdiction in which the copyright-relevant acts underpinning the training of those general-purpose AI models take place. This is necessary to ensure a level playing field among providers of general-purpose AI models where no provider should be able to gain a competitive advantage in the Union market by applying lower copyright standards than those provided in the Union».

logic applies in reverse when Swiss content is used for training abroad. However, this logic may not be accepted without contradiction. We can certainly imagine that if content is collected in Switzerland or some other country, there is a sufficient connection to that country to apply its copyright law to downstream processes (such as use for training). However, this does not correspond to the traditional understanding of the *lex loci protectionis* applicable to determining the applicable law.

- **Contract law**: Agreements made when procuring content for training are also legal requirements. In addition to non-disclosure agreements (see «Protection of secrets» above), these include, in particular, licence agreements and agreements that may restrict the use of content even if it is not protected by law. Such agreements come in a wide variety of forms, e.g. as terms of use when registering on an online platform, when registering for database access or as part of a business relationship in which data is generated. Clauses that specifically restrict the use of content for machine learning are still rather rare. However, they will become increasingly common. If machine learning is not specifically addressed, it must be interpreted in each case whether, for example, a provision that permits the use of content «for internal purposes only» is violated by the use for training large language models. If the use of content is restricted to the «purposes of contract performance», which is standard today, particularly in confidentiality clauses, the use for unrelated training of language models, for example, is likely to be more difficult. If a contracting party nevertheless uses the content for this purpose, it is in breach of the contract and may be liable for damages. In this case, the contract can usually also be terminated as per the terms, or a demand can be made to stop the use that is in breach of the contract. However, a breach of contract is not in itself unlawful; further elements must be added, such as the disclosure of trade secrets, copyright infringement or data protection violations. The latter naturally apply wherever the contract has created the basis for the legally compliant use of content (e.g. under copyright law) by granting a licence to use it.

[10] In practice, someone who wants to train a large language model with certain content must therefore check all these points in relation to this content. They apply in parallel, i.e. the use of content can be unproblematic in terms of copyright but violate data protection and *vice versa*. We have provided a checklist for this purpose at the end of this article.

[11] In the following, we describe approaches for complying with some of the above-mentioned legal requirements. What we do not address in this article is the question of what precautions need to be taken in order to meet the quality expectations for the training of an AI model, some of which may also be of a legal nature. For example, the EU AI Act requires that certain data governance has been implemented as part of the training, validation and testing of AI models for high-risk AI systems in order to meet certain quality requirements. The other obligations for providers of general-purpose AI models under the EU AI Act, for example in the area of documentation, are also not dealt with here. They will apply from 2 August 2025.

V. Copyright law

A. Preliminary remarks

[12] Copyright law is primarily relevant to the training of large language models and the question of whether this is permissible under intellectual property law. This gives the rights holder the exclusive right to determine whether, when and how their work is used (Art. 10 para. 1 CopA). Even if public domain content that is not subject to copyright is also used to train large language models, many training data sets will contain copyright-protected content.[12] This applies, in particular, if a large amount of data from the public internet is used. Anyone who wants to use it to train language models must therefore be able to demonstrate why they is not infringing these rights. There are three possible approaches here:

- The training of large language models with copyright-protected content is no use of the work that is relevant for copyright;
- the use of the content is permitted by law, i.e. there is an applicable exception to copyright; or
- the use of the content is permitted by the rights holder.

[13] We will discuss all three approaches here. The following diagram provides an overview (Figure 1):

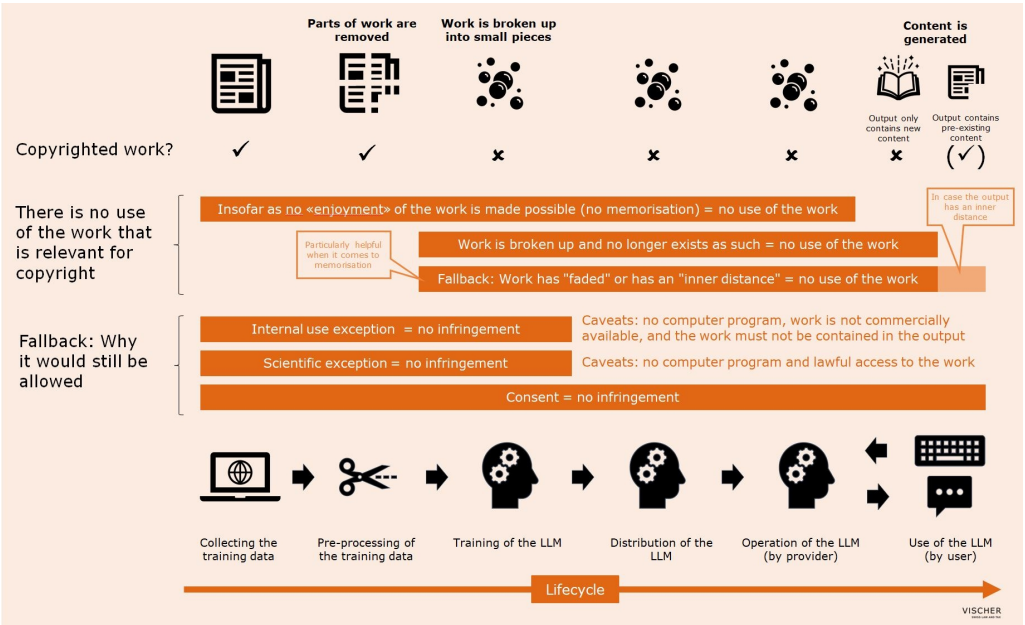

Figure 1: Training a large language model under Swiss Copyright Law

PDF Figure 1

12 Data collections as such do not enjoy any special copyright protection in Switzerland (unlike in the EU). Copyright protection requires that the design or selection criteria of the data collection has an individual character. See also: Supreme Court of the Canton of Zürich of September 1, 1992, in: SMI 1993, p. 331 et seqq., p. 334; Städeli/Mary (fn 7), p. 247 et seq.

[14] With the first two approaches, we are largely entering uncharted territory, at least as far as the training of language models is concerned.

[15] However, before we come to the legal considerations, certain factual points need to be clarified, namely the «problem» of memorisation.

B. Does memorisation take place or not?

[16] In connection with the copyright permissibility of training of a large language model, two constellations must be distinguished: *Firstly*, the constellation in which the training of the model does not lead to any memorisation of the training content, and *secondly*, the constellation in which memorisation of the training content takes place in the model itself. Such a distinction is already important because the memorisation of the knowledge from a series of training content in the model itself (we also refer to this below as the machine-readable representation of knowledge) could be qualified as copyright-relevant copies and the passing on of the model could consequently be qualified as distribution or making available. If, on the other hand, no memorisation takes place, the question of the copyright qualification of the machine-readable representation of knowledge is irrelevant, since there is simply no (relevant) copy of any protected works contained in the model. There are also differences at the level of the use of the model or the output. What the two constellations have in common, however, is that the training itself involves duplication and further processing of the training content.

[17] We will return to this distinction again and again in the following explanations. The focus here is clearly on training a large language model, at least without verbatim memorisation. This is because effective measures are already being taken to prevent such memorisation. However, it can be assumed that memorisation cannot be completely prevented, neither verbatim (as with well-known or frequently occurring slogans) nor – and certainly not – in the area of *content* memorisation. For example, a large language model may not have memorised the sentences from Harry Potter word by word (because appropriate countermeasures have been taken), but it knows the world of Harry Potter and the characters, places and other characteristic features of the Harry Potter universe that appear in it, so that it can easily create a story that takes place in it, if additional measures (e.g. a corresponding so-called alignment,[13] prompt filter or output filter) do not prevent this. We will therefore also discuss the situation of memorisation from a copyright perspective in detail and show how to deal with it.

C. There is no use of the work that is relevant for copyright

1. Approach 1: No «enjoyment» of the work by a human is made possible

[18] As we have already seen, the training of large language models requires adaptations and, above all, (short-term) copies of the relevant training content. Thus, when training language models, an infringement of the exclusive right under copyright law and, in particular, an infringement of the right to produce copies of the work pursuant to Art. 10 para. 2(a) CopA is

[13] For example, the model is taught not to follow prompts.

questionable.[14] In order to answer this question, it is decisive whether it is even possible to speak of a copyright-relevant use of a work in this context.[15] To the best of our knowledge, Swiss court practice has not yet had the opportunity to comment on this.[16] It must therefore be determined by interpretation and on the basis of copyright principles whether the copying processes in the context of training large language models would qualify as infringements of the right to produce copies of the work. Only if this is the case would it be necessary to establish the applicability of an exception provision or the consent of the rights holder to ensure the lawfulness of training with copyright-protected content.

[19] According to Art. 10 para. 2(a) CopA, the rights holder has the exclusive right to produce copies of the work, such as printed matter, phonograms, audio-visual fixations or data carriers. Despite this seemingly analogous wording, copyright law is technologically neutral[17] and the term «copy» is broadly defined.[18] Irrespective of the number of copies of the work produced, it includes both physical and non-physical, temporary and permanent copies that directly or indirectly serve the purpose of consuming the work, i.e. the enjoyment of the work by a human.[19]

[20] At first glance, the qualification of the copies required for the training of language models as copies under copyright law seems obvious (also) due to the wording of the law.[20] Systematic considerations also partly support this view: With the introduction of Art. 24a CopA, the legislator has clarified that copies that are only temporary can also constitute copyrightable acts (otherwise Art. 24a CopA would be superfluous). The copies required for the training of language models do not differ from the copies under Art. 24a CopA for the time being: they are also only of a temporary nature and enable the human consumption of the work in theory (at least if the subsequent use of the language model is included, for which, however, other bodies may be responsible[21]). In principle, this would argue in favour of also qualifying the reproduction processes during the training of a language model as copying under copyright law.

[21] However, there are several arguments against such an understanding. As a first step, it should be noted that the terms of the CopA are to be interpreted independently. Just as not every use of a work constitutes a use within the meaning of Art. 10 para. 1 CopA,[22] not every process that is a reproduction from a purely technical point of view simply qualifies as a copy within the meaning of copyright law. From a technical point of view, there is no doubt that copying takes

14 Sandra Marmy-Brändli/Isabelle Oehri, Das Training künstlicher Intelligenz, in: sic! 2023, p. 655 et seqq., p. 657; Mathis Berger, Künstliche Intelligenz und Immaterialgüterrecht, in: Jusletter IT of July 4, 2024, SN 7 et seqq.; Philippe Gilliéron, Intelligence artificielle: la titularité des données, in: RSDA 2021, p. 435 et seqq., p. 449.

15 This question has also been raised in recent literature: Florent Thouvenin/Peter G. Picht, AI & IP: Empfehlungen für Rechtsetzung, Rechtsanwendung und Forschung zu den Herausforderungen an den Schnittstellen von Artificial Intelligence (AI) und Intellectual Property (IP), in: sic! 2023, p. 507 et seqq., p. 517; Ivan Cherpillod, Intelligence artificielle et droit d'auteur, in: sic! 2023, p. 445 et seqq., p. 446 et seq.

16 Marmy-Brändli/Oehri (fn 14), p. 657.

17 Reto M. Hilty, Urheberrecht, 2nd ed., Bern 2020, SN 304.

18 Manfred Rehbinder/Lorenz Haas/Kai-Peter Uhlig, Orell Füssli Kommentar Urheberrechtsgesetz mit weiteren Erlassen und internationalen Abkommen, 4th ed., Zurich 2022, CopA 24a SN 1; Marmy-Brändli/Oehri (fn 14), p. 658.

19 See among others: Rehbinder/Haas/Uhlig (fn 18), CopA 10 SN 9; Berger (fn 14), SN 8.

20 Thouvenin/Picht (fn 15), p. 517.

21 See: chapter V.E.

22 Denis Barrelet/Willi Egloff, Das neue Urheberrecht – Kommentar zum Bundesgesetz über das Urheberrecht und verwandte Schutzrechte, 4th ed., Bern 2020, CopA 10 SN 8; Marmy-Brändli/Oehri (fn 14), p. 658; Rehbinder/Haas/Uhlig (fn 18), CopA 10 SN 3.

place during the training of large language models. However, according to the above, it is not yet clear that these are copies within the meaning of copyright law.[23] Rather, this must be determined by interpretation.

[22] In the context of this interpretation, it is of central importance that the acts of use reserved to the rights holder in Art. 10 para. 2 lit(a)-(f) CopA do not start with the actual enjoyment of the work by a human – which is free under copyright law – but describe acts that are necessary in the sense of a communication of the work in order to ultimately enable the enjoyment of the work (by third parties).[24] However, when training a large language model, the copies of the training content (and the associated adaptations) are not in a suitable format to enable the enjoyment of this content by third parties, not even indirectly through the use of the model (i.e., the content appearing later in the output). This is the case if no memorisation takes place. However, this is also true in constellations involving memorisation: The content used for training is not available in a form that is suitable for human use, nor can it be perceived by humans, even in the model. Moreover, only the combination with a prompt generates the output.[25] Thus, even in the case of memorisation, the copies of the training content neither directly nor indirectly serve the human enjoyment of works and are fundamentally not suitable for enabling this. Such enjoyment is not the goal of the training either. Rather, these copies are formed into a training data set as part of the training, which is used solely for machine learning thus for higher-level statistical insights. This also applies to the crawler data collections, which are often used as a source and still contain reasonably readable content. However, they are also prepared in a form that is not suitable, let alone intended, for human use.[26]

[23] According to the above, the reproductions of the training content required for the training of large language models do technically constitute copies. However, because the copyright assessment of the processes follows a copyright concept and the (technical) copies of the training content do not enable human enjoyment of the work (not even indirectly), which would be a prerequisite for the copyright concept of use, according to this approach, they are not copies within the meaning of copyright law. Copies of the training content created for the training of a large language model are therefore irrelevant under copyright law.

2. Approach 2: The training is a (free) «enjoyment» of the work by the AI

[24] A second approach draws an analogy with the human consumption of works. The human consumption is – as already mentioned – undisputedly free of copyright. If a person reads a website, for example, reading is not yet a copyright-relevant act that requires the consent of the rights holder; from a copyright perspective, human reading and learning is possible without the consent of the rights holder.[27] The training of large language models with possibly copyrighted content does not really differ from this situation on closer inspection: Similar to the human brain,

23 See also: Thouvenin/Picht (fn 15), p. 517.

24 Federal Supreme Court No. 143 II 617, consid. 5.3.2; Hilty (fn 17), SN 292; Ansgar Kaiser, Der fehlende Werkgenuss beim Text und Data Mining, in: Florent Thouvenin et al. (eds.), Kreation Innovation Märkte, 2024, p. 251 et seqq., p. 253; Rehbinder/Haas/Uhlig (fn 18), CopA 10 SN 3.

25 See: chapter V.C.3.c) for details.

26 An example from CommonCrawl, available at: https://vischerlnk.com/3Z0gmwd.

27 Here still in accordance with: David Bomhard, Text und Data Mining auf Grundlage von Webcrawling und Webscraping, in: InTeR 2023, p. 174 et seqq., p. 175 (explanations refer to German law).

which also has to listen to and read countless texts before it is able to create its own texts, the language model requires training with (copyright-protected) content in order to learn the meanings of the tokens and their relationships to each other.[28] However, in contrast to the human situation, machine reading and learning requires technical duplication of the training content. This does not represent a «fundamental'[29] difference to the human situation that would justify different treatment, because these reproductions are technically conditioned and thus inherent to the enjoyment of the work by the large language model. As things stand today, the enjoyment of works by large language models is only conceivable with such technical reproductions.

[25] For these reasons, too, it can be argued that the (technical) reproduction (and the associated editing) of training content in the context of training large language models does not qualify as copying within the meaning of copyright law and therefore does not infringe copyright. What output is generated with a trained model and what it is used for is – as with humans – a separate issue.

3. Approach 3: AI's «knowledge» lacks copyright relevance

a. Preliminary remarks

[26] In the case of verbatim memorisation of existing works, the question arises as to whether such memorisation qualifies as a copying under copyright law and the passing on of the model consequently as distribution or making available. In addition to the few cases of verbatim memorisation, it is also questionable how to deal with the fact that content from an existing work is incorporated into the linguistic and factual knowledge of the model in such a way that (a) it is present there in a concretisation that is considered a machine-readable representation of knowledge and (b) it can be found in the output in a form in which the individual character of the original work is still recognisable and has not faded due to a suitable prompt in the context of a concrete use.[30] This is not very likely, but it can occur.[31]

b. The works contained in the training data are no longer in the model

[27] The answer to these questions depends first of all on whether the works contained in the training data are continuously retained in the model during the training and thereafter, i.e. during the distribution, operation and use of the model, in a form that falls within the respective scope of protection of the original works. This is already questionable because the works – as already mentioned several times – are broken up into their tokens even in the case of verbatim memorisation.

[28] It is important to recall the legal situation in the analogue world: There, the owner under property law is permitted to split a painting or a photograph, for example, into pieces (unless

28 Thouvenin/Picht (fn 15), p. 517.

29 This is in contrast to: Bomhard (fn 27), p. 175, who sees the fundamental difference to human enjoyment of a work in the duplication of training data required for technical reasons and therefore always requires consent for the training of an AI.

30 The term «fading» is to be understood in the sense of copyright law, i.e., that a work recedes into the background as part of a larger work.

31 Whether the creator of the model can actually be held responsible for this use can be left open at this point. We go into this in detail in chapter V.E.

it is an original work within the meaning of Art. 15 para. 1 CopA). By splitting up a painting or a photograph, for example, the work of fine art or the photographic work is destroyed. It no longer exists as such, even if its individual parts still exist. The individual parts can therefore not constitute a copy in the sense of copyright law either individually or in their disordered entirety.

[29] The training of a large language model is the same (also in the case of memorisation): Firstly, by splitting the works contained in the training data into their tokens and secondly, by using these only to adjust the weights of the model and then «throwing them away», the result is a splitting of the work. Metaphorically speaking, the tokens do not remain together, but the «little screws» of the billions of weights scattered across the entire model are turned slightly in one direction or the other based on their content and their relationship to each other. This process leads to a dismantling of the individual works and thus to their destruction. As a result, the works as such no longer exist. Only the information that has proven to be statistically relevant from the huge volume of training data remains, stored in a form that no longer has anything to do with a text. The relationships between the individual information elements are of a purely statistical-mathematical nature, for example that the word «king» has a much closer relationship to «queen» than to «bicycle». Therefore, according to this approach, the machine-readable representation of knowledge cannot be a copy of works within the meaning of copyright law. It is irrelevant under copyright law.[32]

[30] The fact that works contained in the training data may reappear in the output in the event of memorisation does not change this. In this case, a copyright-relevant reproduction may exist. However, the reassembly of the individual parts or tokens into the pre-existing work constitutes a new generation or «recreation» and not a retrieval of a work stored in the model. The permissibility of such recreation under copyright law is discussed below.[33]

c. Theory of «fading» or «inner distance»

[31] In the event that this view is not followed or the view is taken that the works contained in the training data can be memorised in the language model in a form that potentially falls within the scope of protection of the pre-existing works even after they have been broken up into their individual parts or tokens, the following must nevertheless be considered: The content of the pre-existing works is stored in the model in a form that cannot be perceived by humans, or at least in which a human can no longer recognise the individual work with its individual character, and only the combination with a prompt generates the output. Each piece of memorised con-

32 Other opinion: Tim W. Dornis/Sebastian Stober, Urheberrecht und Training generativer KI-Modelle, Technologische und juristische Grundlagen, in: Roland Broemel et al. (eds.), Recht und Digitalisierung, vol. 19, Baden-Baden, 2024, p. 71 et seqq.: In their expert opinion commissioned by the Initiative Urheberrecht (which represents the interests of rights holders), the authors conclude that it does not matter what actually happens in a model. They argue that it does not matter whether a reproduction is intended or not (p. 75), but overlook the fact that a reproduction can logically only exist if a copy is created at the other end («*re*production»- making one into several), which even in their view is probably not the case here. Instead, their argumentation is based on the fact that the training and use of a model are regarded as a single unit (p. 73), which is not the case. They further argue that only a physical definition of the work is required that is capable of making the work directly or indirectly perceptible to the human senses in some way (p. 78). This argument overlooks the fact that this requires a prompt, which is not contained in the model. A prompt is not only an access key to the otherwise complete «work», but a content-related ingredient, comparable to a two-component glue, which only becomes a glue when both components are mixed. Technically, the building blocks for the work are to a certain extent present in the model, but without the prompt they do not come together.

33 See: chapter V.E.

tent disappears, so to speak, into a «soup» of billions of parameters (weights) and is formed with all other content into a whole (the model) in which the individual character of the individual work is no longer recognisable or has faded. This fading leads beyond the scope of copyright protection. The fact that individual character components of the «soup» can be determined using mathematical methods and the necessary additional aids such as a corresponding prompt (specifically: an output can be generated that falls within the scope of protection of a pre-existing work) does not change the fact that the individual works or their characteristic features fade into the background in the «soup» (here: in the model).[34] The creation of such a «soup» from countless individual works is also not a relevant process under copyright law.[35] The creation and distribution of the «soup» are therefore copyright free, unless the necessary copies are not covered by this exemption in advance or the entire training would be considered a new form of use within the meaning of Art. 10 para 1 CopA. The latter is already questionable because, as shown, it does not serve the human enjoyment of the work itself, but rather to make the work a data point in a statistical collection.

Figure 2: Theory of «fading» or «soup theory»

[32] And even if this «soup theory» is not followed, a copyright infringement due to the verbatim or content memorisation also drops out because there is a sufficient «inner distance» between the pre-existing work and its memorisation. The theory of inner distance states that copyright-

34 A comparison with a photographic mosaic can help here (https://en.wikipedia.org/wiki/Photographic_mosaic): It consists of numerous fully reproduced individual images, which together form a new overall picture. Its subordinate elements, whose individual character is no longer important, lose their copyright relevance in this use.

35 This lends itself to a comparison with the analogous situation in which pieces of numerous picture puzzles (all of which show a protected work) are thrown into a pot and mixed together. The comparison is flawed in that in a language model, a statistical average of all similar pieces is used first. This means that the original images can no longer be fully reconstructed. It is therefore not like shredded paper, which could theoretically always be reassembled.

free use can exist even if the individual character of the original work does not fade, provided that the second work expressly engages with the original work used and thus creates the «inner distance».[36] The secondary work must be considered independent in its essence. A parody or a book review can be examples of such secondary works, the use of which does not depend on the consent of the rights holder of the original work (whereby parodies have additionally been granted an exception provision by the legislator).

[33] Such an inner distance is present here: The purpose of the training of a language model is not to copy the training content, but to determine its meanings, statements and contexts and to record these in a machine-readable form that is independent of the work. It can therefore be compared to the linguist and literary scientist who researches the world of Harry Potter and analyses, for example, how the characters are designed, how they interact with each other, which concepts are used in the stories – and, of course, which linguistic elements are used for this. To do this, he will capture, record and analyse these elements in a structured way. This is exactly what happens when training a large language model. Of course, the linguist's notes can be used to create a new Harry Potter story. Anyone who does this without the consent of the copyright holder or legal permission may be committing a copyright infringement. However, this does not change the permissibility of the scientist's notes under copyright law: they are sufficiently distant from the work and are therefore outside the scope of copyright protection. Similarly, the training of a language model on the basis of protected works is also exempt from copyright.[37]

[34] The machine-readable representation of knowledge is therefore unproblematic from a copyright perspective.

d. Legal consequence of the lack of relevance

[35] We have already explained that Art. 10 CopA describes the actions that are necessary in the sense of the communication of a work in order to ultimately enable the human enjoyment

36 Supreme Court of the Canton of Zürich of July 7, 2009, consid. IV.2.1, in: sic! 2010, p. 889 et seqq., p. 892; Marco Handle, Der urheberrechtliche Schutz der Idee, in: Manfred Rehbinder/Reto M. Hilty/Cyrill P. Rigamonti (eds.), SMI – Schriften zum Medien- und Immaterialgüterrecht, vol. 100, Bern 2013, SN 322 et seqq.

37 This approach can also «solve» another problem, namely the fact that the parameters of a language model can contain a machine-readable representation of knowledge of a work (e.g., a book) without ever having seen it in training. This can be the case with well-known works such as the Harry Potter stories: There is so much material on this in secondary literature, in newspaper reports and on websites that knowledge of the world and stories of Harry Potter will inevitably be assimilated. This is correct because it represents general knowledge. The way language models work means that the model brings this distributed knowledge together like pieces of a mosaic or jigsaw puzzle at certain points in its «soup» to form an overall picture and has this overall picture at its disposal, just like people who have read, seen and heard a lot about Harry Potter and have an overall picture of the world and the story of Harry Potter in their heads without ever having read one of the books. Thus, if the model were asked to do so and no countermeasures were taken, it would also be able to create a story about Harry Potter from it, which could infringe the copyright of the copyright holder of Harry Potter. In this case, however, the copyright «problem» is not the use of the templates (assuming that these are not themselves protected by copyright and do not infringe the rights of the copyright holder in Harry Potter because they have the necessary inner distance, can be based on the right to quote or otherwise fall under an exception to protection), but the result of the training in the form of the machine-readable representation of general knowledge about Harry Potter. In accordance with the explanations above, however, it can be said that in this machine-readable representation the individual features of the original works fade or the machine-readable representation is the product of the model's examination of the training content and thus has the necessary inner distance to fall outside the scope of copyright protection. Furthermore, it can of course also be argued that the training content is not protected by copyright in relation to a specific work (here: secondary literature in relation to the rights to Harry Potter), in which case even a physical copy of this content would not be relevant under copyright law. Only a new creation of an analogous story to Harry Potter could fall within the scope of protection of the original work. It would therefore be purely a question of the permissibility of the output.

of the work (by third parties). The question therefore arises as to which «work» the copies and adaptations enable enjoyment of exactly. On closer inspection, the actions involved in the training of a large language model enable at most the enjoyment of the machine-readable representation of knowledge of training contents. The training content as such cannot be affected by this, because it can only be memorised in the model in the form of the machine-readable representation, and not as knowledge from a single piece of content, rather from a multitude of such pieces of content – in other words, a synthesis. Especially since the machine-readable representation of such knowledge is copyright-free according to what has been said, this must logically also apply to the copies and adaptations that are created or take place during training.

D. Training as a copyright-relevant but permitted act

1. Relevant copyright exceptions

[36] Another approach to solving the problem of training large language models is the application of a suitable copyright exception. We will limit ourselves here to analysing those exceptions that actually come into question. They are the following two:

- **The internal use exception** (Art. 19 para. 1(c) CopA): This provision permits the reproduction of copies of works in companies for internal information or documentation purposes. The complete or substantial copying of a work obtainable commercially is not permitted.[38] However, complete or substantial copies are permitted if they are made when retrieving works that have lawfully been made accessible, e.g. when downloading such content, whereby only the first intended copy made with the download is covered.[39] In principle, the use of the work under Art. 19 para. 1(c) CopA is subject to a remuneration to the author whereby claims for remuneration may only be asserted by the authorised collective rights management organisations. This remuneration already ensures a certain degree of protection for the rights holder, which permits us to more generously interpret the internal use exception. This must always be kept in mind in the following. If the reproduction (i.e. the storage) is made with the intention of generating an output that is (also) directed to third parties outside the company, the reproduction at first glance no longer only serves the purpose of internal information and documentation.[40] However, it can be argued that every internal use fulfils a certain commercial purpose. Even if content is primarily used to inform and educate employees, it is generally intended to have (at least) a positive influence on business operations.[41] If, for example, the employees of a law firm are provided with relevant specialist knowledge in the form of excerpts from academic literature, this

38 For further exceptions, see below.

39 Rehbinder/Haas/Uhlig (fn 18), CopA 19 SN 45.

40 Marmy-Brändli/Oehri (fn 14), p. 659 et seq.; see also: Philip Kübler, Wie generative KI-Systeme Rechte nutzen, medialex 05/23, June 6, 2023, SN 12, available at: https://medialex.ch/2023/06/06/wie-generative-ki-systeme-rechte-nutzen: because the purpose of machine copying in the training of a generative AI system is not internally oriented information but externally oriented production, an AI system cannot claim the purpose of internal information or documentation for itself.

41 See, to that effect: Daniel Schönberger, Deep Copyright: Up- and Downstream Questions Related to Artificial Intelligence (AI) and Machine Learning (ML), p. 18, available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3098315.

is usually done with the aim of keeping employees up to date and thus enabling them to provide competent and prompt advice to clients. The employees analyse and process the information received in order to generate a corresponding output.[42] The Commercial Court of the Canton of Zurich also appears to follow this view in a recent – and criticized[43] – decision, in which it expressly states that the purpose of the information does not have to be limited to the transfer of knowledge (for employees) itself, but may also merely serve to make work easier.[44] With regard to the training of a language model, it can be argued that the copyright-relevant act, namely the copying, is actually only carried out for internal purposes, namely to impart linguistic and factual knowledge to the model. The output that the model later produces is only information derived from the content, which is not itself protected by copyright. Similar to what a lawyer would do, for example, after researching a topic in the internal library in order to prepare an informed response to the client, the model creates an output based on the information processed in this way when it is later used. Here too, however, the problem of memorisation arises: If the work originally used for the training is still recognisable in the output and is thus made accessible to the public, the scope of this exception is exceeded.[45] Furthermore, the internal use exception does not apply fully where content is used for the training of a language model that can still be obtained commercially for a fee and is nevertheless copied completely or substantially. As already mentioned, it would still be permissible to make a copy where such content is legally accessible online. However, one copy is usually not enough, as the content cannot be processed «in one go», but there are numerous intermediate steps with corresponding copies. One possibility would be that only part of such content is used in the first round of training (e.g. 50%) and then the rest in the second round, as the exception to internal use only covers substantial or complete copies.[46] From a functional point of view, even the use of 90% of a work could be distinguished from a substantially complete copy if the 90% refer to each sentence or paragraph, as provided for in certain anti-memorisation techniques such as the *Goldfish Loss method*[47]. No one would refrain from buying a book if they were instead given a copy with 10% of each sentence or paragraph missing, especially as the copy would generally no longer be reasonably usable. In any case, the internal use exception does not apply to computer programs. This is of some practical relevance, as computer programs are also read in in large quantities when training large language models. Finally, the internal use exception also does not permit the reproduction of works of art or the reproduction of musical scores. The internal use exemption is therefore subject to various relevant restrictions and as a result is only of limited use for the training of language models.

42 See also: Elias Mühlemann/Nicole Ritter, Part 10: Copyright and AI: Responsibility of providers and users, available at: https://www.vischer.com/en/knowledge/blog/part-10-copyright-and-ai-responsibility-of-providers-and-users/.

43 Rehbinder/Haas/Uhlig (fn 18), CopA 19 SN 30.

44 Commercial Court of the Canton of Zurich No. HG190187-O of September 6, 2021, consid. 4.3.5.

45 In this sense also: Mühlemann/Ritter (fn 42).

46 According to Federal Supreme Court No. 133 III 473, consid. 3.1 and Barrelet/Egloff (fn 22), CopA 19 SN 30, a substantial copy exists if the purchase of the complete copy becomes uninteresting for an average consumer in view of the scope of the copy. This is the case at the latest when the largest part of the copyrighted material has been reproduced.

47 See: https://arxiv.org/abs/2406.10209.

- **The scientific exception** (Art. 24d CopA): This exception permits the (even complete) reproduction of content, provided that it serves purposes of scientific research and is due to the use of a technical process, i.e. the achievement of the scientific purpose requires the use of a technical process, which in turn requires reproduction. According to the doctrine, this refers to things such as the preparation of data, reformatting and scans with text recognition or labelling.[48] No remuneration is owed.[49] However, access to the content must be lawful, i.e. any fees due (e.g. via registration) must be paid.[50] EU law has a similar regulation under the title «Text and Data Mining» (**TDM**), but it is narrower and gives the rights holder an opt-out right in the case of commercial use.[51] Swiss law (to date) does not provide for this; a corresponding opt-out by a rights holder would therefore not be effective in Switzerland.[52] The doctrine on whether contractual prohibitions of use or conditions are effective in this context is controversial.[53] According to the view expressed here, they are irrelevant insofar as they affect the core content of Art. 24d CopA, i.e. deprive the exception of its effect: The purpose of the scientific exception is precisely to prevent dependence on the consent of the rights holder for reproductions for research purposes,[54] and it cannot be that this purpose can be overridden by a sentence in registration or licence conditions. Moreover, the wording itself only requires lawful *access* to the work – whether the licence granted lapses again is irrelevant. However, the key question regarding Art. 24d CopA is to what extent the training of a language model can be considered a covered purpose of scientific research at all.[55] First of all, it is clear that both non-commercial and commercial scientific research is covered.[56] Scientific research is defined as the systematic, methodical search for new knowledge, regardless of the discipline.[57] The scientific purpose must be clear and specific and be the dominant purpose.[58] The scientific exception therefore also legitimises projects that combine research and commercial applications; if the systematic training of the system serves product development and the knowledge gained is generalised for further product developments, it is irrelevant if a main objective of the project is the subsequent commer-

48 Rehbinder/Haas/Uhlig (fn 18), CopA 24d SN 4; Barrelet/Egloff (fn 22), CopA 24d SN 6.

49 BBl 2018, p. 603 and p. 628.

50 Rehbinder/Haas/Uhlig (fn 18), CopA 24d SN 2.

51 Art. 3 et seq. of Directive (EU) 2019/790.

52 See for a detailed comparison of the different regulations: Damian Hartmann, Text and Data Mining and Copyright in Switzerland and the European Union, in: sic! 2023, p. 157 et seqq., p. 165 et seq.

53 In favour of the mandatory nature of Art. 24d CopA and thus negating the effectiveness of contractual prohibitions of use or conditions: Michael Isler, Text und Data Mining in der medizinischen Forschung, in: LSR 2022, p. 111 et seqq., p. 116; Barrelet/Egloff (fn 22), preamble to CopA 19–28 SN 6; Marmy-Brändli/Oehri (fn 14), p. 664; Michael Isler, Wissenschaftsschranke (Art. 24d URG), in: Peter Mosimann (ed.), Das revidierte Urheberrecht, Bern 2020, SN 230; Melanie Graf/Kirsten Johanna Schmidt, Data Mining und wissenschaftliche Forschung – de lege lata und de lege ferenda, in: sui-generis 2017, p. 185 et seqq., p. 199 et seq.; see also: Berger (fn 14), SN 13; Hilty (fn 17), SN 442 et seq., but not specifically with reference to Art. 24d CopA.

54 See: BBl 2018, p. 602.

55 See, to that effect: Kübler (fn 40), SN 16 et seqq.; Städeli/Mary (fn 7), p. 249; Berger (fn 14), SN 13.

56 See among others: Michèle Burnier, Révision du droit d'auteur suisse: questions choisies sous l'angle du Big data, in: PI – Propriété intellectuelle vol. 13, p. 275 et seqq., p. 286; Rehbinder/Haas/Uhlig (fn 18), CopA 24d SN 7; Isler, Wissenschaftsschranke (fn 53), SN 232.

57 BBl 2018, p. 628; Hilty (fn 17), SN 514.

58 BBl 2018, p. 603 and p. 628 et seq.; Marmy-Brändli/Oehri (fn 14), p. 66; Isler, Text und Data Mining (fn 53), p. 117.

cialization of the application.[59] Nevertheless, the doctrine takes the view that training for commercial applications of artificial intelligence is not a scientific purpose,[60] although this is not really justified and is also incorrect in this generalisation. An AI model is by definition an «insight» because it embodies the result of its analysis of the training content in the form of its parameters (weights) and provides information about how language is used in this training content on the basis of generalised (only machine-interpretable) rules and patterns. This in turn forms the basis for the generation of any new texts with the same language. This insight is undoubtedly achieved both systematically and methodically. The fact that this insight (i.e. the language model) can still be sold later does not change anything, because commercial research is also covered. This means that the training of large language models is covered by Art. 24d CopA. There is now also a first (albeit not legally binding) ruling from Germany, in which even the offering of a data set for the training of AI models was deemed to be a reproduction for purposes of scientific research within the meaning of the TDM regulation applicable there, although the data set was not yet associated with any gain in knowledge, but was only input for training.[61] A restriction also applies here which is that the scientific exception does not apply to computer programs, as well.

[37] Both exception provisions only permit the production of copies without the consent of the rights holder, but not other alterations and adaptations.[62] According to Art. 11 para. 1(a) CopA, the rights holder alone has the right to determine whether, when and how their work may be altered. This raises the question of whether the alterations and adaptations made during the preparation of a training course are alterations of the work in the sense of copyright law and would therefore be prohibited without the consent of the rights holder. In particular, we are talking about reformatting (e.g. conversion into word fragments or tokens or the removal of formatting) as well as redaction (e.g. removal of personal data). According to the view expressed here, these alterations do not constitute alterations of the work within the meaning of Art. 11 CopA nor are they already covered by the reproduction process: *Firstly*, they occur with every reproduction in the digital domain. If a text is scanned and processed with a text recognition system, it is broken up into its individual elements (down to the pixel level), whereby the respective elements are processed individually and distracting elements are hidden for image recognition. Although the content of the work is the same inside the system, it is displayed completely differently. The right of reproduction would lose all meaning if such subordinate, ultimately technically induced processing were not included. This is at least implicitly recognised in the doctrine: In the case of the scientific exception, for example, it describes processes such as the preparation of data, reformatting, scans with text recognition or labelling of data as covered, indeed as necessary for the exception to apply. *Secondly*, there is no copyright infringement if a transformation by a third party causes the individual character of the pre-existing work to fade.[63] The alterations and adaptations in preparation for the training have the effect that the copyright-protected training content is broken up into its parts (or tokens). Copyright protection can also apply to individual

59 Isler, Text and Data Mining (fn 53), p. 117.

60 Rehbinder/Haas/Uhlig (fn 18), CopA 24d SN 6; similarly: Berger (fn 14), SN 13.

61 Hamburg Regional Court No. 310 O 227/23 of September 27, 2024 (not yet legally effective).

62 The internal use within the company also includes the right of distribution and the right to make available within the company: Federal Supreme Court No. 133 III 478.

63 Hilty (fn 17), SN 376 et seqq.

parts of a work if these parts have an individual character (Art. 2 para. 4 CopA). However, the smaller these parts are, the less individual character they have. A single token does not have an individual character, which also leads to the conclusion that the aforementioned alterations and adaptations must be permissible.[64] *Thirdly*, the interests of the rights holder are in no way affected by such alterations and adaptations, which is why, according to the *ratio legis* of copyright law, the rights holders should not have an exclusive right in this respect. Against this background, the preparation of content to be used only for the purposes of training does not require a separate exception provision if the training or the reproductions required for it are covered by such a provision.

2. Consent of the rights holder

[38] Another approach, the consent of the rights holder, is only likely to apply to some of the training content. At first sight, there would appear to be few difficulties in cases where rights are expressly granted. In practice, however, a distinction must be made between the licence granted by the provider of a data set and the licence granted by the rights holder to the content contained therein. From a copyright perspective, only the licence or permission of the latter is usually required, but not the former, because the compilation of a training data set rarely leads to independent copyright protection due to a lack of individuality, which is what makes corresponding licencing necessary in the first place. In the case of content crawled from public websites, however, express consent will hardly ever be given.[65] The question therefore arises as to whether there is usually at least implicit consent to the use of the content. At least in the case of content made freely accessible on the internet by the rights holder, this can be justified with good reason: Anyone who makes content available on the internet today must assume that it will be read not only by humans, but also by machines and used for search engines or AI training.[66] If the rights holder refrains from taking measures against such reasonably expected use, then this indicates implicit consent to this use. Of course, this argument only works as long as the rights holder himself does not (expressly) state otherwise, for example via terms of use, warning letters or built-in, machine-readable coding that prohibits crawling or scraping according to corresponding standards, and of course only where content has been published on the internet with the rights holder's permission.[67]

[39] Another limit to implicit consent is likely to be the memorisation of content: The rights holder will sometimes publish their content for the very purpose of others learning from it or being inspired by it, which is why it can easily be argued that machines are allowed to do this at least as much as humans, at least if they handle what they have learned or been inspired by in the same way. Machines may apply what they have learned or allow the inspiration to take effect, but may not reproduce it or them 1:1 without being authorised to do so by the rights holder or

64 Similarly: Berger (fn 14), SN 15 et seq.

65 Furthermore, due to the extremely large number of works required for the training and testing of AI systems, it is practically impossible to identify and contact the relevant rights holders and ask for their consent to use the works. See: Thouvenin/Picht (fn 15), p. 517.

66 Of the same opinion: Marmy-Brändli/Oehri (fn 14), p. 663.

67 Marmy-Brändli/Oehri (fn 14), p. 663 fn 80.

law, as in the case of the right to quote, for example.[68] From the point of view of consent, it is therefore not the memorisation of the content itself that is problematic, but the generation of output with identically reproduced and thus copyright-infringing content, in particular without complying with the regulations of the right to quote third party words (which language models do not usually do automatically). Of course, this process is not attributable to the person training the model.[69]

E. What if an output infringes copyright?

[40] At this point, the separation between knowledge in the language model (contained in the parameters) and in an output of the model is important: An output is only created by entering a suitable prompt. If this prompt never occurs in the use of the model, the model will never generate such an output. However, the prompt entered is dependent on the user and their use of the model. It cannot be held against the person who has trained the model, at least with regard to the training of the model, especially since at this «level» there are still two mandatory components for an infringement, a prompt and an output. Moreover, the infringement would only be reflected in the output, i.e. it does not affect the model itself. It is like a two-component glue – each component on its own does not stick and is not an adhesive; they only do so in combination. The question of how likely it is that a user of the model will enter a certain prompt, does not affect the model but rather its use and the joint responsibility of the person who makes it possible (more on this below).

[41] In addition, even if a model could be used to generate copyright-infringing output, not every user of such a model automatically commits copyright infringement. For example, if a company uses a general-purpose language model, which has also been trained with Harry Potter content, for an industrial application, and the industrial application never produces a corresponding prompt, this user will not be committing copyright infringement. Even if such a model were to be distributed and copied in its entirety, this would still not constitute an act relevant to copyright law with regard to the output that this language model could generate. In other words, distributing a large language model that can quote Harry Potter is not in itself a violation of the copyright of Harry Potter, because such an act would only manifest itself in the output, i.e. in the specific use of the model.

[42] However, even if a company uses a large language model for the creation and use of an output that infringes third-party copyrights, the person who provides the model to the company is not automatically jointly responsible for this. There are two fundamentally separate actions here: That of training the model and that of using it. Another question that has not yet been raised or discussed in the doctrine is whether the person who provides a model to a user, which

68 It could be argued that certain works are only for entertainment purposes and therefore the analogy with the machine fails because it has no emotions. However, this overlooks the fact that even where a work is primarily for entertainment, this is never an exclusive purpose, but a rights holder always allows secondary purposes such as content analysis. The author of a novel may write it primarily for entertainment, but he is just as happy for the book critic in the arts section, the teacher with his class or the literary scientist interested in certain genres to analyse it with regard to the language used. Several such analyses have been published on various Harry Potter volumes, for example. Of course, due to their inner distance from the original work, these do not constitute an infringement of the same, even if they reproduce the work outside of the right to quote in the context of their analysis.

69 See: chapter V.E.

is then used to commit a copyright infringement, is to be regarded as a *participant* or even an accomplice in this infringement. The provision or passing on of the model could be regarded as a contribution or as an aid to copyright infringement.

[43] In principle, we do not wish to answer this question here, especially since – as we have just explained – it does not directly concern the training of a large language model. However, we would like to point out that this question has already been discussed in court in connection with the responsibility of internet hosting providers for copyright infringements by their customers. For example, we refer to the *Rapidshare case law*, in which a provider whose servers were used for the distribution of pirated copies was found not to have acted as an (commercial) accomplice because its business model was not solely geared towards aiding and abetting copyright infringement.[70] Furthermore, the Federal Supreme Court ruled in connection with a copyright infringement that not every arbitrary act of participation that merely has a «somehow» promoting influence, but is not sufficiently closely connected to the act itself, constitutes a contribution to the infringement.[71] This decision concerned an internet access provider who provided the technical infrastructure for access to the worldwide internet, which was held not sufficient for participation in a copyright infringement of unknown third parties in the context of civil law claims.[72] The provider of a language model who has trained the model himself, but does not operate it himself and only provides it under the condition of copyright-compliant use, will be able to argue that he too only provides a technical infrastructure that does not infringe copyright in itself. This will only be the case if it is used in a corresponding manner (i.e. with a suitable prompt and unauthorised use of the output). The provider of a language model will be closer to the infringer in the causal chain than the internet access provider, but less close than the hosting provider, at least as long as he does not operate the model on his computer and offer it as a service.

[44] The provider case law in the area of personality rights cannot simply be applied to the area of copyright infringements. However, this is mainly due to the fact that in the former case, according to prevailing practice, any (even subconscious) form of «contribution» to an infringement of personality rights can be prosecuted independently under civil law in accordance with Art. 28 para. 1 SCC, provided, however, that there is a causal connection.[73] In a case concerning the registry of IP addresses, the Supreme Court of the Canton of Solothurn even considered that participation was sufficient and that an adequate causal connection was not required in order not to undermine the low requirements for «participation»,[74] which, of course, misses the point because adequacy is a figure of liability and criminal law, which was not at issue in the case in question.

[45] In connection with internet search engines, the District Court of the city of Zurich came to the conclusion that the entry of search terms in a search engine is not attributable to the operator, at least insofar as the search terms are not suggested. It did not consider the display of hits by the operator of the search engine to be a violation of personality rights.[75] Previously, in another

70 For further information: https://steigerlegal.ch/2022/03/07/rapidshare-urteil-volltext/, archived at: https://perma.cc/8UC2-TBT8.

71 Federal Supreme Court No. 145 III 72, consid. 2.3.1.

72 Ibid., consid. 2.3.2.

73 Federal Supreme Court No. 5A_792/2011 of January 14, 2013, consid. 6.2 et seq.; Federal Supreme Court No. 141 III 513, consid. 5.3.1; for the delimitation see: Federal Supreme Court No. 5A_658/2014 of May 6, 2015, consid. 4.2.

74 Supreme Court of the Canton of Solothurn No. ZKBER.2022.17 of November 3, 2022, consid. 5 et seqq.

75 District Court of the city of Zurich No. CG190002 of October 26, 2020, consid. 2.1.3.

decision, the same court had attributed a contribution to the operator because it had made information available online to a wide range of users that would not otherwise have been able to find it.[76]

[46] In a new decision concerning FIFA and Google, the Commercial Court of the Canton of Zurich went further: the operation of a search engine and the associated processing of information was not sufficient for the court to grant passive legitimation in the case of allegedly infringing articles. The articles at issue were only displayed on the first page of search hits in the search engine when search terms containing words from the title of the articles were entered. The court found that it was therefore only the user of the search engine who established the link between the articles and FIFA. Google would make the search engine available and thus facilitate the discovery of articles. However, this is not yet a legally relevant act of cooperation because it is not sufficiently closely connected to the act itself; a mere «effect» is not sufficient, «cooperation» is required. According to the court, Google's conduct that – in addition to operating the search engine – specifically enabled or facilitated the finding of the articles had not been alleged.[77]

[47] This case law can be applied with good reason to the providers of large language models and general-purpose AI chatbots based on them, such as «ChatGPT» or «Copilot»: search engine operators record all kinds of content from the internet in their databases, index it and thus make it searchable. Providers of AI chatbots record the same content, except that they do not map the content 1:1 in their models, but only aggregated linguistic and factual knowledge derived from it. This in turn is made available for free retrieval as part of a service. With regard to content that infringes any third-party rights but is not reproduced (because measures have been taken to prevent verbatim memorisation), they go one step further, as they do not redistribute or make such content accessible 1:1. Even in the case of memorised content, they only provide building blocks that they compile for the user according to their prompt, like a search engine that finds suitable content in its index and provides access to it. And those who merely offer copies of the models they have trained (e.g. in the form of files for installation on the user's own computer) go one step further – comparable to those who provide the raw index to the operator of a search engine.

[48] If a user of such an AI chatbot enters prompts that generate an unlawful output, this is not to be assessed differently than entering specific search terms in a search engine that leads to unlawful articles. Similarly, the mere provision of the AI chatbot (or even just the model) cannot yet be considered a relevant contribution to the infringement, at least if a specific prompt is required for the output in question, which users do not simply enter, and which is not suggested by the provider. If this is already the case in personality rights, where any contribution is sufficient, this must apply all the more in the area of copyright law, as there are higher requirements for relevant contributions to the infringement. The liability of providers of general-purpose AI chatbots and (even more so) the AI models on which they are based for the outputs generated by their users with special prompts is therefore severely restricted, at least under Swiss law. This appears to be correct: an AI model can generate any number of different outputs. In combination with an appropriate prompt, these can always result in infringement. However, to the extent that the user intends this and the provider of the chat bot or even the model does not specifically encourage

76 District Court of the city of Zurich No. CG160047 of June 1, 2018, consid. 6.2.9.

77 Commercial Court of the canton of Zurich No. HG220030-O of August 21, 2024, consid. 3.2.4.2.6.

this,[78] the necessary adequate causal connection[79] or «cooperation» is ultimately lacking because large language models have a very broad range of possible applications. For the sake of completeness, the case in which an AI service provider feeds the language model used by it, together with the user's prompt, with further information (e.g. current information, content from databases or from the internet) in order to optimise the output, is to be distinguished from this. This so-called «Retrieval Augmented Generation» (RAG) and the copyright permissibility of using third-party works in the input or prompt of a language model must be assessed separately; this is not the issue here.

F. Excursus: forum shopping

[49] As a result of the principle of territoriality in international intellectual property law, works protected by copyright are protected in each country in a limited territory in accordance with the respective national copyright law.[80] As a result, this «domestic» copyright can only be infringed by an act committed at least partially within the country.[81] In other words, the place of the allegedly copyright-infringing use is decisive for the localisation under substantive law.[82]

[50] In the area of training large language models, the relevant act of use is the storage of the training content in the AI provider's training corpus and the actual training of the AI system, and not its subsequent use by end customers. Accordingly, the applicable law depends on where the AI provider carries out the training of the large language model.[83] Empirically, this can be demonstrated by the fact that AI training is usually carried out in countries with «liberal» copyright regulations and far-reaching TDM exceptions.[84]

[51] AI providers therefore have the option of seeking out for a legal system that is favourable to them because it is «liberal» for the training of large language models – even if the AI system trained in this way is subsequently placed on the market in Switzerland or the EU, for example.[85] If this is a legal system that exempts the training of large language models with copyright-protected content, for example, there is nothing to prevent this language model from subsequently being placed on the market in Switzerland or the EU, at least from a copyright perspective. Moreover, as noted above, this is relevant under the AI Act.[86]

78 E.g., with the operation of a model or AI system specifically for unlawful purposes or training to promote infringing outputs.

79 There is always a natural causal relationship, because without a model there is no output, just as there are no search hits without a search engine index.

80 Gerhard Schricker/Ulrich Loewenheim, Urheberrecht, 6th ed., Munich 2020, preamble to UrhG 120 SN 109.

81 Schricker/Loewenheim (fn 80), preamble to UrhG 120 SN 126, SN 131 et seqq. and SN 142 et seqq.

82 Niklas Maamar, Urheberrechtliche Fragen beim Einsatz von generativen KI-Systemen, in: ZUM 2023, p. 481 et seqq., p. 486.

83 For example: Maamar (fn 82), p. 486.

84 Marcus von Weiser, Generative KI und Urheberrechtsschranken, in: GRUR-Prax 2023, p. 516 et seqq., p. 520 SN 39; Dornis/Stober (fn 32), p. 120.

85 See for the whole and with reference to the EU: Maamar (fn 82), p. 486.

86 See: para. 9 above.

G. Conclusion

[52] When training large language models, content that is protected by copyright is used alongside content that is in the public domain. However, training with copyright-protected content is already in principle legally *permissible de lege lata* – this applies both to training with and without memorisation. This can initially be argued on the grounds that the reproduction of works for the training of a large language model does not constitute a copyright-relevant act, as these processes do not enable human enjoyment of the individual training contents, which is a prerequisite. If this is rejected, it can be argued that the training is to be seen as an «enjoyment» of the work of the AI, which must be copyright-free analogous to the human enjoyment of the work because it corresponds to it in relevant aspects. The technically required reproductions are inherent to this «enjoyment» and therefore do not constitute a significant difference to the human enjoyment of the work.

[53] Training with memorisation is more «problematic» in that, compared to training without memorisation, a machine-readable representation of knowledge of training contents is added, which in turn must be checked for copyright admissibility. To create this machine-readable representation, however, the individual works are broken up into their individual parts or tokens. As a result of this process, the works contained in the training data are destroyed and no longer exist as such. For this reason alone, the creation of the machine-readable representation cannot constitute a copyright-relevant act.

[54] If this approach is not followed and the machine-readable representation of knowledge of training contents is nevertheless considered to have something like the quality of a work and thus a potential ability to infringe rights, the theory of «fading» or «inner distance» should also be pointed out: In the case of the machine-readable representation of knowledge, the individual characteristics of the copyrighted training content sufficiently fade or the model creates a sufficient inner distance to this content, which means that the machine-readable representation is not covered by the scope of the copyright protection of the training content. The consideration of the model alone is therefore not at all suitable for enabling the training content to be perceived as a work, which is why the same applies to the reproductions, alterations and adaptations required in the context of the training as well as to the transfer of the model to a third party, as applies to the training without memorisation. However, because the output of the model can constitute a copyright infringement due to memorisation, training with memorisation raises the further question of the responsibility of the person who trains the model and makes it available for use. In principle, we do not want to discuss this question conclusively at this point. However, it can be said that – analogous to the situation with hosting providers and search engine providers, where there is already corresponding case law – there are good reasons for rejecting liability in principle, at least for those who train a general-purpose model and offer it as such to the general public.

[55] If an act relevant to copyright law is nevertheless assumed, it can be argued that permission has been granted by the rights holder, at least for certain works. Where this has not been done expressly on the basis of licences, it has been done implicitly through the public provision of content on the internet without restrictions. In addition, two copyright exceptions can be cited: Firstly, the internal use exception and secondly, the scientific exception. Both have relevant restrictions (e.g. they do not apply to computer programs, which is often relevant when training language models). Nevertheless, they appear to be practicable where there is lawful access to the training content (scientific exception) or it is ensured that the relevant works do not appear in the output (internal use exception).

[56] Finally, AI providers can use *forum shopping* to apply a «liberal» legal system for the training of large language models which exempts such training with copyrighted content, for example. From a copyright perspective, there is then nothing to prevent the large language model from being placed on the market in Switzerland at a later date because the act of training has been completed. The same applies if it is to be distributed in another legal system that also follows the *lex loci protectionis*.

[57] Our comments show that an amendment of Swiss copyright law is not mandatory. However, it is also to be expected that the use of content for training AI models will remain a controversial political issue due to the actual or perceived economic interests at stake: one side will demand that Switzerland becomes (more) attractive as a location for the development of AI models, while the other side will demand protection of its content from exploitation, particularly by US technology companies. It remains unclear whether this can really be achieved by adjusting copyright law since the law regularly lags behind reality in such cases, insofar as it can effectively counter the power of the factual. Finally, there will also be those who will demand that Swiss copyright law be aligned with that of the EU, particularly with regard to the TDM regulation – for example, by introducing an opt-out right within the framework of the scientific exception. However, this raises additional questions. Furthermore, Switzerland has not done badly so far by being cautious about regulating technology (the Federal Council's report on the need for regulation in the field of artificial intelligence was not yet available at the time of writing).

VI. Data protection law

A. Preliminary remarks

[58] For the training of a large language model to be admissible under data protection law, various conditions must be met according to the Data Protection Act (**DPA**), which in turn depend on whether the controller is a private person or a federal body (e.g. a public research or educational institution).

[59] The following conditions apply to private persons:

- the processing principles of Art. 6 DPA must be observed;
- if this is not the case or if there is an objection, justification is required (Art. 30 para. 2(a) or para. 2(b) DPA in conjunction with Art. 31 para. 1 DPA);
- the duty to provide information pursuant to Art. 19 et seqq. DPA must be complied with;
- any international transfers must be made in accordance with Art. 16 et seqq. DPA; and
- if sensitive personal data is disclosed to third parties, justification is also required (Art. 30 para. 2(c) DPA in conjunction with Art. 31 para. 1 DPA).

[60] As far as a federal body is concerned (in particular the research institutes and the Federal Institutes of Technology such as ETH Zurich), the following conditions apply:

- the processing principles of Art. 6 DPA must be complied with, unless the law provides otherwise;
- the duty to provide information pursuant to Art. 19 DPA must be complied with;
- a legal basis is required (Art. 34 para. 1 DPA);

- a statutory basis in a formal law is required in the following cases:
  i. the matter involves the processing of sensitive personal data,
  ii. the matter involves profiling, or
  iii. the purpose or manner of the data processing may lead to a serious violation of the data subject's fundamental rights (Art. 34 para. 2 DPA);
- any cross-border disclosure of personal data must comply with Art. 16 et seqq. DPA.

[61] Whether these requirements are met must be assessed depending on the source of the personal data. It makes sense to make the following differentiations and assumptions:

- A distinction must be made between **public sources** and non-public sources. If a public source contains personal data, this does not mean that the data has been made public with the knowledge and will of the person concerned. In most cases, this will not be the case.[87] However, the distinction is important in order to assess the interference with the data subjects' personality rights that is associated with the use of the source for training. This in turn is important for the assessment of proportionality, possible justification and the need for a statutory basis in a formal law. This last point should be put into perspective insofar as, where the publication constitutes a particularly serious interference and the published data is not easily accessible (e.g. publication of stolen private data on the darknet), the interference caused by the use of this data will be correspondingly serious. Conversely, the hurdles to the use of data where it has been published with the knowledge and will of the person concerned will pose fewer problems under data protection law.[88] In the present case, we assume that only public sources will be used for the training. The sources qualify as public sources if they are accessible to an indefinite number of people, even if a fee has to be paid or a contract concluded.
- **The memorisation of personal data** is likely to be a rare occurrence and therefore only affect very few people. If the sources are public sources, experience shows that only people who are (or have been) widely reported on in public or who have published a lot under their own name (even if they are not otherwise «persons of contemporary historical relevance») are affected. We refer to them as «public figures». Whether memorised personal data is output for later use depends on whether the model is given appropriate prompts.[89]
- We assume that there is a certain interest in the memorisation **of data on public figures**, i.e. that a model can comment on persons of contemporary historical relevance. We do not assume that there is a particular interest in statements about people who are not persons of contemporary historical relevance but who have published a great deal, but we do not consider them to be bothersome either.
- The training of a large language model **does not in itself serve any purposes related to specific persons**. Although it can and will contain personal data of public figures who fre-

87 Luca Dal Molin/Kirsten Wesiak-Schmid, Datenschutz im Unternehmen, Zurich/St.Gallen 2023, § 1 SN 202, cite the example of published photos on the internet in which third parties are depicted against their knowledge or the publication of personal data in a newspaper against the will of the persons concerned.

88 Corrado Rampini/Rehana C. Harasgama, Basler Kommentar Datenschutzgesetz/Öffentlichkeitsgesetz, 4th ed., Basel 2024, DPA 30 SN 26.

89 See also: Rosenthal, Part 19 (fn 1).

quently appear in the training data, this is not the primary goal. Rather, the goal is the creation of a system with linguistic and factual knowledge (general knowledge) that can generate suitable output based on an input.[90] However, where the general knowledge also includes knowledge about specific people of contemporary historical relevance, its memorisation is a secondary goal.

Recently, experts have even gone one step further and argued that the processing of data for the purpose of training a language model is not the processing of personal data at all, because the entity carrying out the training would not be able to identify the data subjects with the effort it is willing to make due to the circumstances.[91] We are not discussing this theory for the time being; however, it could also be advocated under Swiss law.

- Training is a **self-contained processing procedure**. It includes the preparation of the training data, the execution of the training (not supervised or monitored) and the validation or testing of the model. For our purposes, we assume that the training data is archived, but this is not relevant for the language model. The use of the language model or its further processing is not part of this processing operation.
- In principle, **no illegal or immoral sources** such as data from the darknet or from websites with hate speech are used for the training. No data will be used from sources that, at the time of collection, have objected to the use of their data for training purposes (e.g. by means of anti-crawling indicators) by means of recognised standards.[92]
- We assume that the resulting large language model will be made **available to** the **public** (i.e. it will not be used solely for internal purposes and remain confidential). However, the source material will not be made publicly accessible.

[62] In addition to the substantive requirements of data protection law, we believe that a **data protection impact assessment** must be carried out as a formal requirement for training. Training a language model also constitutes a processing activity that must be recorded in a **register of processing activities**. Whether **logging** (namely the storage and deletion of personal data) and **processing regulations** are also required depends on whether a federal body is involved and whether sensitive personal data is processed on a large scale (which is unlikely).

B. Compliance with the processing principles and the obligation to provide information

[63] The processing principles and the obligation to provide information are complied with as follows:

- **Principle of purpose limitation:** The collection of training data for training generally represents an indirect collection of personal data for a new purpose: The data is not collected from the data subject themselves (exception: a private individual's own website is imported)

90 See also: David Rosenthal, Part 17: What is inside an AI model and how it works, available at: https://www.vischer.com/en/knowledge/blog/part-17-what-is-inside-an-ai-model-and-how-it-works/.

91 Peter Craddock: Op-ed: AI training data = (non-)personal data? And is consent really relevant?, in: LinkedIn, October 14, 2024, available at: https://vischerlnk.com/3A27ZYm.

92 However, see: excursus in chapter VIII.

and it will generally not have been apparent to them that the data is to be used by the controller for training (exception: the data comes from a data source that has communicated this to the individuals because it wanted to be able to pass on their data for such purposes, e.g. a social media platform that sells data for AI training). The question therefore arises as to whether the training is at least compatible with the purpose for which the data was originally obtained from the data subject. The use of data obtained for a primary purpose for a secondary purpose not related to specific persons in pseudonymised or anonymised form is considered a typical example of a purpose that is «compatible» with the original purpose.[93] This is normally the case here: If a language model is trained with content that contains personal data, this is not normally represented in the model, i.e. the model only remembers aggregated information about how language is formed. This results in anonymisation. An exception to this is the case where a certain personal date (e.g. the birthday of a well-known person) occurs sufficiently often in the training data and is therefore included in the model's language knowledge. In this case, pseudonymisation takes place because the persons concerned are not directly identified on the basis of the language knowledge, but rather a suitable prompt is required to make them recognisable (the information cannot be recognised at all on the basis of the parameters in the model, which is why some people assume that a language model per se cannot contain any personal data[94]; however, this view – which we consider incorrect – has not really been accepted[95]).[96] In these cases, the compatibility with the purpose may be called into question. Yet, the following counterargument can be made: If the high frequency is due to the fact that the training data reflects public content, then it will generally be personal data of public figures, because by its very nature, personal data can only be found in large numbers and in different public content in the case of public figures (with or against their will). However, if this is the case, then such a person must now expect that this information about them will also be used for the training of large language models, because it is now common knowledge that as much public content as possible is used for this purpose. In any case, it can be argued with good reason that it is not «unexpected, inappropriate or objectionable» (to quote the words of the Federal Council's legislative commentary),[97] that such personal data is not only used for numerous publications, but also for the training of large language models. Thus, the principle of purpose limitation (Art. 6 para. 3 DPA) is basically fulfilled for both memorised and non-memorised personal data in training data.

- **Principle of proportionality**: The processing of personal data must be proportionate in accordance with Art. 6 para. 2 DPA, i.e. suitable, necessary and reasonable for the pur-

93 David Rosenthal, Datenschutz beim Einsatz generativer künstlicher Intelligenz, in: Jusletter of November 6, 2023, SN 28.

94 Hamburg Commissioner for Data Protection and Freedom of Information, Discussion Paper: Large Language Models and Personal Data, July 15, 2024, available at: https://datenschutz-hamburg.de/news/hamburger-thesen-zum-personenbezug-in-large-language-models.

95 European Data Protection Board, Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, December 17, 2024, available at: https://www.edpb.europa.eu/system/files/2024-12/edpb_opinion_202428_ai-models_en.pdf.

96 David Rosenthal, How and why a large language model can know the «birthday» of a (public) person, illustration from August 4, 2024, available at: https://www.rosenthal.ch/downloads/VISCHER_LLM_birthday_example.pdf.

97 BBl 2017, p. 7025.

pose of the training.[98] This principle is usually based on the fact that the training of a large language model depends on it containing as much training data as possible, i.e. the quality improves as the data volume increases.[99] This is where caution must be exercised as what is important is not the largest possible volume of personal data, but rather the training data. Training with personal data is not the goal. However, personal data is often unavoidable in the training data, for example when public articles, essays, websites or entries from Wikipedia are used. However, they are still suitable for training: if the model is to be taught how sentences must be formulated in a certain language in order for them to be correct, they must also contain all the elements that the language entails – and this includes information that turns the sentences into personal data, such as names or other identifying elements. In this respect, personal data is also necessary for training. However, it is not necessary for the information to be true, only for there to be exemplary formulations. Training sentences such as «Peter Parker was born on 1 January 1980» are sufficient to convey the concept of the sentence. An exception exists where a higher number of repetitions can be expected, i.e. in the case of public figure. In this case, there is an interest in ensuring that the language model can not only formulate sentences relating to birthdays correctly, but that it also provides the information correctly when it comes to the birthday of a public figure. This means that the model can only use personal data for training in cases where public figures are involved. In other cases, it is not necessary and could be replaced with pseudonyms or redacted without harming the training (to the extent that a model must of course also learn how, for example, a telephone number is used in a language context and therefore also needs sentences where this happens; however, again the telephone number does not have to be a real for this to work).[100] The third aspect of proportionality, reasonableness, on the other hand, poses fewer difficulties: For data subjects whose personal data is not memorised, the interference with their personality rights is only very limited. It is not about them, and their personal data is not in the model. On the other hand, there is the interest of the person training the language model. Their interest is to use as much training data as possible. In doing so, they have only limited possibilities to remove unnecessary personal data in advance with reasonable effort, insofar as such data can be recognised at all. However, if this is done, we believe that it is not only reasonable, but also necessary to process the personal data that remains in the cleaned training data. This is because the purpose of the training can only be fulfilled if it can be carried out with reasonable effort. If the training data had to be checked individually for any personal data, training would no longer be possible from the outset and the purpose would be frustrated. In this sense, the principle of proportionality is also complied with.

- **Principle of fairness, including the principle of transparency:** Processing must continue to be carried out in good faith (Art. 6 para. 2 DPA). In addition to the criteria of proportionality already mentioned, this means that the processing must be carried out in a

98 Lukas Bühlmann/Michael Reinle, Basler Kommentar Datenschutzgesetz/Öffentlichkeitsgesetz, 4th ed., Basel 2024, DPA 6 SN 51 et seqq.; with further references: Dal Molin/Wesiak-Schmid (fn 87), § 1 SN 133.

99 Steffen Albrecht, ChatGPT und andere Computermodelle zur Sprachverarbeitung – Grundlagen, Anwendungspotenziale und mögliche Auswirkungen, TAB-Hintergrundpapier Nr. 26, 2023, available at: https://www.bundestag.de/resource/blob/944148/30b0896f6e49908155fcd01d77f57922/20-18-109-Hintergrundpapier-data.pdf.

100 See: Rosenthal, Part 19 (fn 1).

comprehensible, transparent manner.[101] It means that it must not be generally perceived as offensive, unfair or disruptive.[102] In our opinion, the latter cannot be assumed in the case of the training of a large language model (assuming that the language model serves legitimate private or public interests and not, for example, illegal purposes). If copyright law – as shown – already intends to permit the analysis of third-party works in order to gain knowledge from them (here: language knowledge), this proves that our legal system fundamentally wants to permit such processes and even intends to make them possible, including where the processes interfere with third-party legal spheres – at least as long as these are not significantly affected. This is also the case here, provided that no purpose related to specific persons is being pursued and no memorisation takes place. In our opinion, the above-mentioned exceptions for persons whose data is already frequently in the public domain do not preclude this, as mentioned above, but will normally be covered by a legitimate public or private interest. This will not necessarily apply to data from data breaches or other unlawful data sources, but such data will not be used according to the premises mentioned. The same applies to data from websites that have opted out of crawlers for the purposes of AI training. Under these circumstances, the training of a large language model does not appear to be unfair, offensive or disruptive, even beyond the principle of proportionality. The recognisability of data processing, i.e. its transparency, is somewhat more difficult. Transparency should enable data subjects to object to the use of their data.[103] As the data is not collected directly from the data subjects, it is not possible to inform them directly in most cases. The controller therefore only has the option of providing general information about its procurement and further processing of personal data, for example via corresponding announcements or its privacy notice. The drawback of this approach is that the data subjects have no reason to inquire about the relevant information on the websites in question. However, in our view, the principle of good faith does not require absolute transparency, but rather transparency that appears fair, i.e. in view of the circumstances. These circumstances also include the severity of the interference that the processing of the data entails. As shown, it is not significant because the processing is neither personal nor does it lead to the storage of personal data. In other words, personal data is, in principle, merely incidental. In comparable cases, Swiss law does not normally see any special need for protection, such as the right to one's own image, which also does not apply if persons are merely in the background of a photograph.[104] This is ultimately also the case with the duty to provide information under Art. 19 DPA. Although it also exists in principle in the present case, it ceases to apply if it is not possible to provide the information or providing the information requires disproportionate effort (Art. 20 para. 2 DPA). This is the case here. Nevertheless, we recommend that those who wish to train a large language model should state this accordingly in their privacy notice. In this case, we believe that both the principle of fair processing and the duty to provide information are fulfilled.

101 Bühlmann/Reinle (fn 98), DPA 6 SN 50.

102 See further: Bühlmann/Reinle (fn 98), DPA 6 SN 49.

103 See: David Rosenthal, Part 16: How company can ensure transparency for their use of AI, available at: https://www.vischer.com/en/knowledge/blog/part-16-how-company-can-ensure-transparency-for-their-use-of-ai/.

104 Andreas Meili, Basler Kommentar Zivilgesetzbuch I, 7th ed., Basel 2022, SCC 28 SN 20.

- **Principle of storage limitation:** Personal data may only be stored for as long as it is necessary to fulfil the purpose of processing (Art. 6 para. 4 DPA). A distinction must be made between personal data in the model and personal data in the training data. If memorisation takes place, the personal data naturally shares the fate of the large language model: as long as it is used, any personal data it contains will also be needed. This will hardly pose any difficulties in practice. Rather, the more pressing question arises as to the retention of training data beyond the training itself. However, the question of whether training data should be stored must be considered in a broader context. It may not only serve the actual training, but also other purposes, such as providing evidence of the sources and training data that were actually used (for example in the event of legal disputes, but also for reasons of transparency). They can be relevant for scientific research, for example to investigate aspects of memorisation. They may also be necessary for re-training. Any of these reasons may be sufficient for not deleting training data immediately after training. The principle of storage limitation (a partial aspect of the principle of proportionality) will therefore generally not be violated as long as there is still a good reason to retain the training data.[105] However, access to this data must be sufficiently restricted.
- **Principle of accuracy:** Anyone who processes personal data must ensure that it is accurate, whereby accuracy is determined by the purpose for which the personal data is processed (Art. 6 para. 5 DPA).[106] At first glance, this requirement appears to be impossible to meet, as training requires training data in quantities that cannot be checked in detail. Even with internet crawler data alone, a large amount of information from websites that is factually incorrect must be expected. In our view, however, this is not relevant here, especially as the purpose of training a large language model is not to build up factual knowledge from individual pieces of content. Rather, the purpose is to build language knowledge, i.e. the model should capture how language is used in the training data, and also build up general and specialized knowledge, i.e. the model should capture what information occurs frequently enough that it can be considered to fall into that category. The training is the processing procedure with which these purposes are achieved – and with regard to these purposes, the personal data in the training data is legally correct if it reflects the training content in an unadulterated manner. The purpose is not the creation of any superordinate knowledge, but the knowledge contained in *the respective training content*. If the factual knowledge also relates to public figures, it may also contain their personal data if the information is seen frequently enough during the training. With regard to this purpose, the principle of accuracy does indeed require that appropriate measures are taken to ensure accuracy.[107] However, this does not normally cause any difficulties either: Models do not, as already mentioned several times, include every piece of personal data as such, but only those pieces of information that occur the same way in larger numbers, in other words that are «confirmed» by a wide variety of sources. The date of birth of a public figure, which is mentioned in numerous publications, is such an example, as already mentioned. Since memorisation only takes

105 See: Stephanie Volz, KI Sandboxen für die Schweiz?, in: SZW 2022, p. 51 et seqq., p. 56.

106 Bühlmann/Reinle (fn 98), DPA 6 SN 247; Nadja Braun Binder, Künstliche Intelligenz und automatisierte Entscheidungen in der öffentlichen Verwaltung, in: SJZ 115/2019, p. 467 et seqq., p. 474.

107 Bruno Baeriswyl, Stämpflis Handkommentar Datenschutzgesetz, 2nd ed., Bern 2023, DPA 6 SN 62 et seqq.; Reto Fanger, Orell Füssli Kommentar Datenschutzgesetz, 2nd ed., Zurich 2023, DPA 6 SN 11.

place in such cases, it automatically fulfils the requirement of a fact check. After all, it is necessary to use as wide a variety of training data as possible for training a large language model and to give preference to good quality sources. In this sense, more personal data also serves data protection. The accuracy of the memorised personal data must also be distinguished from the accuracy of the personal data in the output of a language model; this can also contain personal data that has not been memorised as such, but rather represents a hallucination[108] (or originates from the user's input). Since the use of a large language model represents a separate processing activity, this does not affect the training. The principle of correctness is therefore fulfilled under these conditions.

- **Principle of lawfulness:** Finally, personal data must be processed lawfully (Art. 6 para. 1 DPA). This requirement refers to legal violations outside the DPA, such as those that can occur when the betrayal of a trade secret is exploited (Art. 162 SPC).[109] It could be argued that the use of personal data published on the darknet as spoils from a hacker attack also constitutes such unlawful processing. For this reason, for the purposes of training a large language model, care must be taken to ensure that no data from illegal or immoral sources is used, which is, however, common practice anyway. The fact that «normal» websites may occasionally contain leaked trade secrets or illegal content should not cause any difficulties in the present case: Firstly, «unlawful» in the present case only means provisions that serve to protect the personality of the persons concerned (and not other «victims», such as the company whose trade secrets are affected), and secondly, in any case, individual personal data are merely «incidental» as shown, i.e. they are not the point and they only fall into the hands of the controller by chance. In our opinion, under the above premises, they should not lead to data processing being considered unlawful. The principle of lawfulness will therefore also be complied with.
- **Principle of data security:** A level of data security appropriate to the risk must be guaranteed by taking suitable technical and organizational measures (Art. 8 DPA). Security is understood as ensuring the confidentiality, integrity and availability of the data and the traceability of data processing.[110] In the present case, this means that it must be ensured for the storage of the training data and the training processes themselves that the aforementioned protection objectives are not violated. In the case of language models, particular attention must be paid to technology-specific attack methods, such as the risk of «poisoning» a language model with training data that is mixed with non-authentic information in order to introduce it into the language model. It would be conceivable, for example, to introduce certain falsified factual knowledge about a person through sufficiently frequent repetitions in a training data set. The falsified factual knowledge would then be reflected in the model and retrieved when suitable prompts were given. This can be partially counteracted on the one hand by creating the training data sets oneself, and on the other hand by

108 Hallucinations are not a problem of the accuracy of the training or the personal data in the model, because this does not usually refer to the fact that a model contains incorrect information (this is of course possible if there is sufficient «incorrect» training data), but that it does not contain any or sufficiently clear information on a particular topic and therefore invents such information to fill gaps because it wants to provide the user with a complete text. This is therefore not a data protection problem of the model, but of its use.

109 David Rosenthal/Yvonne Jöhri, Handkommentar zum Datenschutzgesetz, Zurich 2008, oDPA 6 SN 3.

110 Christa Stamm-Pfister, Basler Kommentar Datenschutzgesetz/Öffentlichkeitsgesetz, 4th ed., Basel 2024, DPA 8 SN 2 et seq.; see also BBl 2017, p. 7031; Clara-Ann Gordon/Luisa Egli, Orell Füssli Kommentar Datenschutzgesetz, 2nd ed., Zurich 2023, DPA 8 SN 1.

using existing data sets that are known to be reliable. The measures do not have to guarantee perfect data security but must be appropriate to the risk. Based on previous experience, we assume that the principle of data security can also be complied with.

- **Prevention of bias and discrimination:** Not addressed in the above principles is the prevention of bias and discrimination, which is often mentioned in connection with data protection requirements for large language models.[111] These requirements relate to the use of language models and their suitability for certain applications, but not to the training as a process for processing personal data. Insofar as no memorisation of personal data takes place, the data subjects are not affected by any distortions in the model, at least with regard to their personal data used for training. Any other persons whose personal data is used by the model are affected. Those persons whose personal data is memorised may also be affected. In this respect, the problem depends on the data sources used for the language model. Each individual source does not pose a problem in terms of data protection law. However, if a balanced selection of data sources is not made with regard to the purpose of the language model, this can lead to the memorisation of one-sided personal data in relation to these persons and thus to the aforementioned distortions in the model knowledge, which in turn can affect both the principle of accuracy and that of good faith and proportionality, insofar as a use case provides for the retrieval of corresponding data. Attention must therefore be paid to this circumstance, even if it can only affect the admissibility of the training under data protection law insofar as the stated intended use has already been established at this point in time.

[64] The above explanations show that the training of a large language model under the above premises does not in principle lead to a violation of the processing principles of Art. 6 and 8 DPA. Therefore, justification under Art. 31 DPA is generally not required for **private controllers** (but see below).

C. Grounds for justification

1. The controller is a private person

[65] However, justification would be necessary for private controllers if the training would lead to the memorisation of sensitive personal data and it could reasonably be expected that the latter would be retrieved if the language model were used or shared (Art. 30 para. 2(c) DPA).[112] One example would be if the political views of a public figure were memorised in the model (for example, the large language models know the views of Donald Trump). Justification is also required in cases where processing for the purposes of training is objected to (Art. 30 para. 2(b) DPA).[113] This cannot be ruled out and is particularly likely if a corresponding objection is noted on websites or in other sources that originate from the data subject themselves and contain personal data.[114]

111 See on the risk of discrimination: Braun Binder (fn 106), p. 473 et seqq.

112 Rampini/Harasgama (fn 88), DPA 30 SN 20; Thomas Steiner/Christian Laux, in: Adrian Bieri/Julian Powell (eds.), Kommentar zum Schweizerischen Datenschutzgesetz mit weiteren Erlassen, Zurich 2023, DPA 30 SN 23 et seqq.

113 David Rosenthal, Das neue Datenschutzgesetz, in: Jusletter of November 16, 2020, SN 38 et seq.

114 See further: Steiner/Laux (fn 112), DPA 30 SN 14 et seq.

The situation is similar in cases where personal data is collected from a source that has excluded the use of the personal data for the purposes of the training due to its intended purpose.

[66] Possible grounds for justification are the following:

- **Processing for purposes not related to specific persons** (Art. 31 para. 2(e) DPA): This justification suggests an overriding private interest in the training of a large language model provided that the training does not serve the memorisation of personal data. As already mentioned several times, this requirement will usually be met because training serves to build up language and general knowledge, not to record individual personal data. Exceptions exist in relation to public figures whose personal data are memorised, but this justification does not apply to them anyway. The processing of their data must therefore be justified in another way. The justification can only be invoked if these three requirements are met: (1) The data must be anonymised as soon as the purpose of processing permits.[115] This is not a problem in relation to the training, as the aggregation of the information from the training content, to which the training process naturally leads, also results in the necessary anonymisation. The storage of training data, on the other hand, serves other purposes that must be justified separately. (2) As far as possible, sensitive personal data may only be disclosed to third parties in such a manner that the data subject is not identifiable.[116] In practice, this requirement will generally not represent a relevant hurdle either. (3) The results are published in such a manner that the persons concerned are not identifiable.[117] By its very nature, this requirement is also normally met, subject to cases in which memorisation takes place. However, memorisation alone is not sufficient to make them identifiable: the data subjects are only (but nevertheless) identifiable if it can be reasonably expected that a corresponding prompt will be entered that leads to the retrieval of the information in question.[118] The justification for processing for purposes not related to specific persons therefore applies in principle. Excepted are those cases in which personal data is memorised and its retrieval must be expected.
- **Public figures** (Art. 31 para. 2(f) DPA): Furthermore, there is generally an overriding private interest in collecting personal data about a public figure insofar as the personal data relate to that person's public activities. This justification is likely to cover much of the personal data contained in the training data, which originates from those persons who have a higher potential for memorisation. Although the justification only covers the collection of such data, according to the doctrine this also includes related activities such as recording, separating and cataloguing.[119] It is therefore conceivable that training could also be regarded as a type of «cataloguing», especially since it does not *a priori* make a difference whether the collected data is «stored» in a conventional database or a language model. However, passing on the language model would no longer be considered «collecting» if the personal data in question were memorised in it and its retrieval could be expected. This justification therefore no longer covers these cases, which may occur.

115 Rampini/Harasgama (fn 88), DPA 31 SN 63; BBl 2017, p. 7076: A *de facto* anonymisation is sufficient.

116 Rosenthal (fn 113), SN 42; Steiner/Laux (fn 112), DPA 31 SN 33.

117 Monika Pfaffinger, in: Bruno Baeriswyl/Kurt Pärli/Dominika Blonski (eds.), Datenschutzgesetz (DSG), 2nd ed., Bern 2023, DPA 31 SN 87.

118 See: Rosenthal, Part 19 (fn 1).

119 Rosenthal/Jöhri (fn 109), oDPA 13 SN 70.

- **Overriding private and public interest** (Art. 31 para. 1 DPA): Even if none of the justifications listed in Art. 31 para. 2 DPA apply, there may still be an overriding private or public interest in data processing.[120]

  In the present context, this interest lies in being able to build a large language model. This requires the largest possible amount of training data that reflects how language is used. Furthermore, in order to increase the quality of the model, there is an interest in ensuring that this training data is as diverse as possible, which requires the development of as many different sources as possible. In turn, training a language model is only possible if it can be carried out with reasonable effort. The effort required to reliably anonymise all training data would represent an insurmountable hurdle. With the amount of text that is processed, only highly automated processes can be used that do not require a great deal of computational effort, as otherwise the training would already fail at data cleaning. If anonymisation is too extensive, the training data also loses information that is important for correct and balanced training results. If the model is made available to the public, the public also has an interest in ensuring that the model has been trained optimally. Finally, the person who trains a model has an interest in keeping the effort involved as low as possible. As far as personal data of public figures is concerned, depending on the intended use of a language model, the provider and its users may also have an interest, and ultimately also a public and private interest, in knowing these people and being able to make statements about them, for example. These statements should also be as accurate as possible. This in turn presupposes that as much data as possible about a person is processed, firstly because this reduces the probability of hallucinations (i.e. that the model does not have the requested or required information, but considers it necessary and therefore invents it) and secondly because the aggregation of the information read, which naturally occurs during the training of a language model, reduces the weight and therefore the influence of individual pieces of information and therefore also misinformation in the totality of a particular piece of information. If a few articles about a person contain certain false information, this will be given less weight in the language model or even «ignored» if there are many more articles that contradict this false information with correct information.

  On the part of the data subjects, a distinction must be made between those whose data is used for training but not memorised, and those whose data occurs so frequently that it is included in the model as personal data:

  In the first case, the normal case, there is a certain loss of control over what happens to one's own data and for what (external) purpose it is used. However, the use is neither related to specific persons nor does it have a lasting effect: although the personal data is read by the model, it is not incorporated into it as such; it merely contributes to the formation of the model's language knowledge but does not survive this process, i.e. it is not memorised either literally or analogously. If the data is public, the interference with the personality rights of the person concerned is even less. This is because if information is public, it must be expected that the same data will be consumed by a person and will remain in their memory as individual pieces of information, which will not be the case with a language model. The use of such content for training a language model is much more controlled and limited,

120 Rampini/Harasgama (fn 88), DPA 31 SN 27 and SN 69.

and is therefore less serious. Furthermore, any measures taken counteract the loss of control, such as the automated redaction of certain identifying information such as telephone numbers, credit card numbers, social security numbers, addresses and names. However, the reasonably possible effort involved prevents these measures from being perfect. Rather, the aim is to limit the use of personal data overall. The most effective approach is the aggregation of personal data through the training process and the associated anonymisation, which is system dependent. In this normal case of training without memorisation, we therefore believe that the private and public interest in training outweighs this.

In the second case, the exceptional case where a memorisation of content takes place, a distinction must be made as to whether the memorisation is due to the fact that the person in question is a public figure or person of contemporary historical relevance. If this is the case, there will usually be a public interest in the memorisation. Special measures to ensure this will not normally be necessary if the training data records are carefully selected. Content memorisation will only occur if the content in question appears in the same form in a very large number of sources, which in turn will generally only be the case when there is genuine public interest.

However, two exceptions are conceivable:

- A certain piece of personal information occurs so often because the person concerned has frequently made it public themselves (e.g. because content from social media in which the person was very active was used for the training). In this case, it can be assumed on the basis of Art. 30 para. 3 DPA that there is no violation of personality rights because the information was made generally accessible with the knowledge and consent of the data subject[121] and the use by search robots (including those that provide content for AI training) is within the scope of what is expected. Justification or weighting of interests is therefore unnecessary.
- The second case concerns «relative» persons of contemporary historical relevance, who are no longer of public interest today, but whose data still frequently appears in the training content. In this case, content memorisation can occur. However, this does not automatically mean that this personal data will be retrieved. It will only be included in the output if it is specifically searched for using a corresponding prompt, which is unlikely in the case of relative persons of contemporary historical relevance. If this is the case, it can be argued that the language model does not contain this personal data despite memorisation because it is unlikely to be extracted (so-called relative approach).[122] If such a prompt nevertheless occurs, the person concerned may have to defend themselves against the output of their data with the relevant user of the language model, analogous to the current situation with search engines, where the search engines must block a result in the hit list even if it is contained in the original source on the internet and (even legally) freely accessible (and the search engine is also permitted to index the content; only the output must be blocked in the case of certain search queries). This can also be expected of the person concerned. The reverse is true for the person who trains a language model and obtains information from numerous

121 Pfaffinger (fn 117), DPA 30 SN 67.

122 See: Rosenthal, Part 19 (fn 1).

generally accessible sources of information; otherwise, training would be impossible from the outset.

As a result, the training of a large language model is generally justified by an overriding interest. In individual cases, this may not apply without further ado, but the persons concerned can generally be expected not to prohibited training in such cases, but to require the output of the language models in question to be filtered in order to protect their personality rights, which is comparable to the established practice for search engines.

- **Consent:** In the context of Art. 30 et seq. DPA, the question may also arise as to whether the data subject has given their consent or whether the data has been made generally accessible with their knowledge and consent, meaning that there is no violation of personality rights, provided that the data subject has not expressly prohibited processing in the latter case (Art. 30 para. 3 DPA). While these two aspects may be relevant in specific individual cases (e.g. in the case of claims by a specific person), they are not practicable as a basis for justifying training, as they require a case-by-case assessment. However, this is not possible due to the volume of data. Even in the case of data from platforms that primarily contain data published by data subjects themselves (e.g. social media networks), there are likely to be quite a few exceptions, meaning that this basis cannot be considered sufficiently reliable.

2. The controller is a federal body

[67] It should be noted that only private controllers can invoke the aforementioned justifications. In the case of **federal bodies** (or other state bodies that have to act in accordance with cantonal laws), in addition to consent in individual cases and the protection of life and limb, the only justification for cases in which data subjects object to processing or one of the processing principles cannot be complied with is their own **legal basis**[123] which is necessary in any case. This means that in these cases it must be checked whether the legal task justifying the training of the language model or the legal authorisation for data processing, which allows the training of the language model, also covers these cases. This also applies to the disclosure of sensitive personal data, which requires justification in the case of private controllers. Federal bodies do not require a justification, but in the case of sensitive personal data, their legal basis must be a formal law (or one of the exceptions applies, such as the processing of personal data for purposes not related to specific persons in accordance with Art. 39 DPA).[124]

[68] Let us take the legal basis of the Federal Institute of Technology as an example. Art. 36c of the ETH Act (**ETHA**) stipulates that within the scope of research projects, the two federal institutes of technology and the four research institutes within the ETH domain may process personal data, including sensitive personal data, insofar as this is required for the given project. They must ensure that the provisions of the DPA are complied with. This regulation provides a comprehensive legal basis in a formal law. Although the regulation is not very detailed, this is due to the nature of the matter: only those processing operations that are *necessary* for a research project are permitted, even if they involve sensitive personal data. Applied to the present example, the question arises as to whether the training of a large language model can be described as a «research

123 Rampini/Harasgama (fn 88), preliminary remarks to DPA 30–32 SN 1.

124 See the above comments on the parallel standard for private persons as controllers.

project». Here we can refer to our comments above on copyright law, where «research» is also used as the basis for the corresponding limitation provision. As we have already explained there, we believe that the training of a large language model can certainly be described as research: It pursues the goal of gaining knowledge about the use of language from the training data through a machine procedure and is preserving this knowledge in a machine-usable form. When large language models are trained at the ETH, this is also done to answer numerous other questions that arise in the still largely unexplored field of machine learning of large language models. Incidentally, questions relating to data protection issues such as the triggers of memorisation and its prevention are also to be clarified during this process, which in turn requires the use of personal data, as no knowledge can be gained in this area without appropriate training and language models. In our opinion, the fact that certain models are ultimately made publicly accessible does not prevent them from qualifying as research; this is no different from the publication of a scientific paper including the corresponding measurement results. Against this background, we conclude that the legal basis, in this context, covers the training of a large language model. The absence of memorisation is not a prerequisite for this. On the contrary, it can be the subject of research. In these cases, it is not the legal basis that has a corrective effect, but the principle of proportionality and, in particular, proportionality in the narrower sense, in which the interests in and against data processing must be weighed up, analogous to the weighing up of interests in the area of private data processing. The above considerations on justification based on overriding private and public interests can be applied analogously here.

D. Cross-border disclosure of personal data

[69] Finally, we would like to point out that any cross-border disclosure of personal data must comply with the requirements of Art. 16 et seqq. DPA.[125] This will not be a challenge when training large language models in Switzerland, as there will be no such disclosure, even if the training data originates from abroad. The fact that a language model created in this way is later also used cross-border is also not relevant here, as the training of a large language model and its subsequent use represent two different processing activities. However, the question may be relevant if a company carries out its training in a cloud environment located abroad, whereby no new difficulties arise here other than those generally associated with cloud use. However, the Federal Council's adequacy decision with regard to the CH-US Data Privacy Framework (DPF) has eased the situation considerably since September 2024.[126]

E. Conclusion

[70] Under data protection law, the training of a large language model in Switzerland using public content is generally permissible even without the consent of the data subjects. It is important to

[125] Accordingly, personal data may only be transferred to other countries if they have adequate data protection legislation or, if not, if one of the guarantees in Art. 16 para. 2 DPA or one of the exceptions in Art. 17 para. 1 DPA applies. For further information: Christian Kunz, in: Adrian Bieri/Julian Powell (eds.), Kommentar zum Schweizerischen Datenschutzgesetz mit weiteren Erlassen, Zurich 2023, DPA 16 SN 16.

[126] David Rosenthal, Swiss-US DPF: How to transfer data to the US with and without it, August 16, 2024, available at: https://www.vischer.com/en/knowledge/blog/swiss-us-dpf-how-to-transfer-data-to-the-us-with-and-without-it/.

note that the nature of the training of a language model means that the personal data is only included in the language model in aggregated form, which means that the interference with the personality rights of the data subjects is minimal. Therefore, there is normally no requirement for justification. In exceptional cases, however, a justification may be required for private-sector controllers. The justification is usually given as overriding private or public interests. This even applies to memorised personal data, as memorisation is usually due to the fact that the data is found in numerous public sources. Under these circumstances, it can be assumed that there is public and private interest or that the data was made accessible with the knowledge and consent of the persons concerned. In the case of federal bodies, the respective legal basis for their own data processing must also be examined; in the case of the ETH, for example, Art. 36c ETHA provides such a basis for the creation of large language models and the associated further research.

VII. Unfair competition law

A. Scope of the UCA

[71] The UCA is applicable to acts of competition.[127] This includes acts which are objectively designed to influence competitive conditions and are not carried out in a completely different context, whereby the infringer's conduct must be market-relevant, market-oriented or competition-oriented.[128] Neither a subjective intention to influence competition nor an actual influence on competition is required.[129] A competitive relationship with the affected suppliers or customers is also not required.[130] In contrast, actions in a purely private context or activities that only have an effect within the company are excluded from the scope of the UCA.[131]

[72] The training of a large language model therefore does not fall within the material scope of the UCA if it serves purely private, non-commercial purposes or only has internal company effects. This is the case, for example, if the language model to be trained is only used within the company and only in such a way that it does not influence the competitive position in external relationships (e.g. by improving competitiveness due to the ability to complete orders faster, better or cheaper than the competition); internal experiments with AI are an example of this. In contrast, outside of the purely private or only company-internal area the scope of the UCA is wide: Accessing and using the freely available content for training free of charge constitutes an act of competition because the person training the model thus saves himself the effort that would have been necessary to originally generate this training content.[132] Such an act is objectively capable of influencing the competitive situation.

127 Reto M. Hilty, Basler Kommentar UWG, Basel 2013, UCA 1 SN 33.

128 Federal Supreme Court No. 120 II 76, consid. 3a.

129 Eugen Marbach/Patrik Ducrey/Gregor Wild, Immaterialgüter- und Wettbewerbsrecht, 4th ed., Bern 2017, SN 1126.

130 Reto Heizmann, Orell Füssli Kommentar Wettbewerbsrecht II, 2nd ed., Zurich 2021, UCA 1 SN 26.

131 Marbach/Ducrey/Wild (fn 129), SN 1127; Heizmann (fn 130), UCA 1 SN 26.

132 See this idea in another context: Federal Council's legislative commentary on the UCA 1983, p. 1047 et seqq., p. 1070; Federal Supreme Court No. 4C.342/2005 of January 11, 2006, consid. 3.2.

B. In particular Art. 5 UCA

[73] The admissibility of the training of a large language model under unfair competition law depends primarily on the provisions of Art. 5 UCA.[133] Accordingly, a person acts unfairly in particular if they (i) exploit a work product entrusted to them, such as an offer, calculation or plan, without authorisation; (ii) exploit a third party's work product, such as an offer, calculation or plan, even though they must know that it was given or made accessible to them without authorisation; and (iii) adopt and exploit another person's work product that is ready for the market by means of technical reproduction processes without any reasonable effort of their own.

1. Exploiting of a work product?

[74] All three cases under Art. 5 UCA involve the exploitation of a (third-party) work product. Work products are products of intellectual effort and material expenditure that are not protected outside the scope of special legislation on the protection of intellectual property.[134] Mere ideas, flashes of inspiration and unspecified methods do not constitute work products, even if they are set down in writing or materialised.[135] Exploitation is understood to mean any economic utilisation of another person's work products.[136]

[75] It is questionable whether the training of a large language model with the corresponding content represents an exploitation of work products:

- In terms of learning language knowledge, this is to be rejected on the grounds that, in the case of the work products, the product of intellectual effort is generally not the language used in the content. Rather, the language is the means by which the actual work product, namely the factual knowledge embodied in the content, is expressed or materialised. It could be argued that language and thus language knowledge can also be a product of intellectual effort – namely when language is seen not as a means to an end, but as an end in itself («linguistic art»). However, this is likely to be statistically rare and thus not «seen» often enough by the model, so that this language knowledge is not included in the model.
- When it comes to learning factual knowledge, it should be noted that the training is not about taking over the specific product of the intellectual effort, but rather about selecting individual pieces of information contained in it, merely in the sense of data points to form an overall statement. The input into the process is therefore different from the output, and only the output is exploited. It is comparable to a person who reads several essays on a topic by different authors and then, in a separate work product, records in their own words the information that seems certain to them after having read them all. They have not exploited any of the preexisting work products but have merely drawn their own conclusions from them. As with language knowledge, factual knowledge is not dependent on the individual content; this only has an effect if and when there is other content that, to a certain extent,

133 In this sense also: Städeli/Mary (fn. 7), p. 249 et seqq.; Philippe Gilliéron, Réflexions autour de la contractualisation des projets d'intelligence artificielle, in: sic! 2024, p. 423 et seqq., p. 425.

134 Federal Supreme Court No. 117 II 199, consid. 2a/ee; Federal Supreme Court No. 122 III 484, consid. 8b.

135 Federal Supreme Court No. 122 III 484, consid. 8b.

136 Federal Council's legislative commentary on the UCA 1983, p. 1069; Markus R. Frick, Basler Kommentar UWG, Basel 2013, UCA 5 SN 53.

says the same thing, i.e. when there is «more of the same». What is adopted is not the individual content, but an aggregate of everything. It is therefore not about the opinion, recommendation, compilation, etc. of the individual on a topic, but rather the «prevailing» doctrine on it. Determining and using it must remain free and should not be monopolised by those who write it down or contribute to it, even under the unfair competition law.

[76] Therefore, in our opinion, it can be argued with good reason that the training of a large language model does not represent an economic utilisation of work products and thus no exploitation of the same.

2. Art. 5(a) and Art. 5(b) UCA

[77] Both case constellations require «entrustment» in one way or another. In the case of so-called direct exploitation (case constellation (i)), this is already clear from the wording of the provision. It is aimed at those situations in which someone has come into possession of the work product in agreement with its creator.[137] In contrast, case constellation (ii) extends the scope of application of Art. 5 UCA to persons who have not been entrusted with the work product directly by the creator, but who have gained possession of it without authorisation in some other way (so-called indirect exploitation).[138] Although the wording of the provision does not expressly mention this, the unanimous opinion is that the work product must have been entrusted by the creator to the person who hands it over to the infringer.[139] Entrustment presupposes a contractual, pre-contractual or quasi-contractual relationship,[140] which results in a prohibition of exploitation at the expense of the person receiving the work result.[141]

[78] Both case constellations (i) and (ii) therefore require at least a contract-like relationship (either between the creator and the infringer in (i) or between the creator and the intermediary in (ii)). A direct contractual, pre-contractual or quasi-contractual relationship between the creator and the potential infringer, i.e. the person who trains the language model, is unlikely to ever exist. However, we do deal with this case constellation, in which received content is used in breach of contract for AI training purposes, in Chapter IX. A contractual or quasi-contractual relationship between the creator and the intermediary is also unlikely to be the rule: content on the internet is very often simply taken from somewhere and redistributed – think of social media and so-called reposts, for example. It is usually a «chain» of information flow, whereby the content is generally only entrusted to the first intermediary by the creator, but not to the intermediary whose content is then effectively used to train the large language model. The criterion of being entrusted therefore already excludes a lot of content from the scope of application of case constellations (i) and (ii).

[79] In case of constellation (ii), the potential infringer, i.e. the third party who exploits the work product for his benefit, must also know that the work product has been made available or acces-

137 Federal Council's legislative commentary on the UCA 1983, p. 1069.

138 Lukas Fahrländer, DIKE UWG Kommentar, Zurich 2018, UCA 5(a) and 5(b) SN 24.

139 Federal Council's legislative commentary on the UCA 1983, p. 1070; Frick (fn 136), UCA 5 SN 58; Fahrländer (fn 138), UCA 5(a) and 5(b) SN 25.

140 Federal Supreme Court No. 133 III 431, consid. 4.5.

141 Frick (fn 136), UCA 5 SN 44.

sible to him without authorisation. In other words, the required subjective component for the potential infringer relates to the breach of the contractual, pre-contractual or quasi-contractual prohibition of exploitation.[142] Negligent ignorance is sufficient for civil liability.[143] However, this does not mean that the potential infringer has a general duty to inquire; such a duty can only be affirmed in circumstances that suggest that the documents provided by the intermediary were not prepared by the intermediary.[144]

[80] In order for the user of the content to be negligent with regard to the unauthorised transfer or unauthorised access, they must violate standard due care. In this context, the considerations regarding provider liability[145] can be applied analogously: the prevailing opinion in legal doctrine is that hosting providers are not obliged to carry out preventive checks on all content; only if they are expressly made aware of a violation of privacy, for example, are they obliged to investigate it and, if necessary, remove the content in question.[146] One reason for this is that the effort for the platform operators would otherwise be enormous and actually not feasible. With regard to internet search engines, the Geneva Tribunal de première instance ruled that it was unreasonable to expect search engine operators to check every single website listed by them for illegal content.[147] The same must apply in the present context: due care cannot be understood to mean that, in the context of training large language models, every single piece of content must be checked to see if there is any indication that the content has been entrusted to the intermediary and is thus being provided or made accessible to the user in violation of a prohibition on exploitation. Otherwise, the training of large language models would be prevented from the outset, which cannot be the telos of Art. 5 UCA or the UCA as such. Rather, the user's good faith is to be assumed as long as he takes appropriate measures to ensure that the data collection activities for training do not include any content from known illegal sources – such as Darknet forums for the distribution of stolen data – and specifically excludes sources where he has been specifically and in a relevant way made aware that they contain content that violates fair trading law. This requirement is likely to be met if, in the context of crawling for AI training purposes, commonly available block lists are used, such as those also used by companies and providers to filter their internet access (to prevent employees or customers from using prohibited or dangerous websites) and certain techniques, such as those necessary to access most darknet websites (e.g. TOR browsers), are not used; in our opinion, a case-by-case assessment or research to create your own block lists is not necessary. Alternatively, content with a corresponding reputation that has already been crawled and is available on the market can be used.

142 Fahrländer (fn 138), UCA 5(a) and 5(b) SN 1.

143 District Court of the city of Zurich No. GG040064/U of August 23, 2005, in: sic! 2006, p. 112 et seqq., consid. XV.3.1; Heizmann (fn 130), UCA 5 SN 15; Frick (fn 136), UCA 5 SN 59; Fahrländer (fn 138), UCA 5(a) and 5(b) SN 27.

144 Frick (fn 136), UCA 5 SN 59.

145 For a comprehensive overview, see for example: Federal Council, Die zivilrechtliche Verantwortlichkeit von Providern, report of December 11, 2015, available at: https://www.bj.admin.ch/bj/de/home/publiservice/publikationen/berichte-gutachten/2015-12-11.html; also: David Rosenthal, Internet-Provider-Haftung – ein Sonderfall?, in: Peter Jung (ed.), Tagungsband Recht aktuell, Edition Weblaw, Bern 2006.

146 See with further reference: Report of the Federal Council (fn 145), p. 63; Alexander Kernen, Volle Verantwortlichkeit des Host Providers für persönlichkeitsverletzende Handlungen seines Kunden, in: Jusletter IT of March 4, 2013, SN 16; Betty-Annett Meier, II. Theoretischer Teil / F. – H., in: Roland Müller/Thomas Geiser/Kurt Pärli (eds.), Bewertung des Arbeitgebers im Internet, 2018, p. 59.

147 Tribunal de première instance No. C/9894/2007-17 of November 4, 2008.

3. Art. 5(c) UCA

[81] In contrast, case constellation (iii) may well exist at first glance: This initially covers all public content that can be considered a work product that is ready for the market (i.e., for example, content from media, social media or platforms for user-generated content).[148] On the one hand, such work products must be adopted and exploited «as such», and this must be done «without any appropriate own effort». If work products are thus sufficiently modified before they are (re)used or if sufficient effort is made, the requirements of this provision are not met.[149] While criteria such as «market readiness» and «work product» will not easily be met, given the amount of content used to train a large language model, a lot of content on the internet will undoubtedly still meet these criteria. It therefore depends on whether a work product is *adopted* as such and without appropriate own effort and – according to the prevailing case law[150] – also *exploited* as such and without reasonable own effort. Exploitation requires that the adopted work product is supplied to the market, because only then does it fall within the scope of protection of the UCA.[151]

[82] None of this can be assumed *prima vista* when training a large language model: Not only is the effort involved in training usually considerable (see below), but training data is not transferred as such, rather the linguistic and factual knowledge it contains is aggregated. This means that exploitation as such is naturally ruled out, at least as long as the work product is not reflected «as such» in the output of the language model. As in the case of copyright, the absence of memorisation also helps to ensure legality here (with the difference that unfair competition law can also apply to content that is not protected by copyright[152]): if memorisation can be prevented, then the adoption of the work product «as such» is already lacking, and even more so its exploitation. If, on the other hand, it occurs, it could be argued that the work product in question (e.g. an article that has been used several times for training) has been incorporated into the model as such. However, with this starting position, it can also be argued that although the content of the work product has been incorporated into the model, it has not been adopted «as such», but in a different form, namely as aggregated linguistic and factual knowledge that can only be used in combination with a suitable prompt to generate a text that more or less corresponds to the original work product. This is because training content is never stored «as such» in the model itself, but in a form abstracted from it, as has already been shown above in connection with copyright.[153] It would then have to be shown that the work product can be retrieved in its original form with the appropriate prompt and that such a prompt can also be expected from a third party in the market in order to assume (and prove) that the work product is exploited as such. These combined circumstances are likely to be rare, if they exist at all (see, however, the US case in which the «New York Times» sued «OpenAI» and «Microsoft»).[154]

148 Rolf H. Weber/Lennart Chrobak, DIKE UWG Kommentar, Zürich 2018, UCA 5 SN 15 and 18; Court of Appeal of the Canton of Bern of May 21, 2001, in: sic! 2001, p. 613 et seqq., consid. 9; Federal Supreme Court No. 131 III 384, consid. 4.2.

149 Heizmann (fn 130), UCA 5 SN 20 and SN 24.

150 Federal Supreme Court No. 131 III 384, consid. 4.3.

151 Weber/Chrobak (fn 148), UCA 5(c) SN 25.

152 Marbach/Ducrey/Wild (fn 129), SN 8; Hilty (fn 17), SN 19 *e contrario*.

153 Chapter V.C et seq.

154 See: https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html.

[83] If memorisation cannot be prevented, at least not in individual cases, the question arises as to whether at least an appropriate own effort is made to adopt or exploit or both. A double effort comparison must be made: on the one hand, the performance of the creator must be compared with that of the potential infringer and, on the other hand, the performance of the potential infringer must be compared with their hypothetical expenditure of effort if they was to follow the individual production steps themselves.[155] An infringement of unfair competition law is only deemed to have occurred if the potential infringer cannot be shown to have made an reasonable effort in either comparison.[156]

[84] In particular, the effort required by the potential infringer must be compared with the effort required by the creator.[157] It should be noted that the training is automated and that not much computing power is required to process a single text. However, the training of a large language model requires a high level of computing power overall, because it relies on being fed with a large amount of text, which must also be curated accordingly. The investment required to carry out training is correspondingly high.[158] While it can be argued that only the fraction of the potential infringer's effort that is attributable to the work product in question should be taken into account for the comparison of effort, in our opinion it can also be argued, and with more convincing reasons, that the high initial investment should be taken into account *in globo* – because there is no question of exploiting the individual work product, but rather only of exploiting all work products together in the form of the large language model. Furthermore, it is correct to first consider the time and effort that the creators have spent on the general language knowledge contained in the content, as this is what is initially important when training a language model. Under these circumstances, the «first» effort comparison will already lead to the assumption that the potential infringer's own effort for adoption and exploitation is appropriate, because the effort that the creator spends on the language skills is usually negligible and individual exceptions are not relevant due to the way in which language models learn. The same will also apply to factual knowledge that is adopted, because this must be so common and widespread that it also occurs in numerous other sources.

[85] Finally, it should be noted that the protection provided by Art. 5(c) UCA is limited in time. This provision only applies as long as the costs of producing the work have not already been amortised.[159] Thus, the time limit alone should largely exclude the application of Art. 5(c) UCA to the training of a large language model.[160]

155 Federal Council's legislative commentary on the UCA 1983, p. 1071; Federal Supreme Court No. 131 III 384, consid. 4.4.1.

156 Federal Supreme Court No. 131 III 384, consid. 4.4.1; see also: Reto Arpagaus, Basler Kommentar UWG, Basel 2013, UCA 5 SN 92.

157 For the training of an AI system see: Städeli/Mary (fn 7), p. 250; in general: Weber/Chrobak (fn 148), UCA 5(c) SN 50.

158 For example, see: Inside IT, So viel kosten grosse KI-Modelle, April 22, 2024, available at: https://www.inside-it.ch/so-viel-kosten-grosse-ki-modelle-20240422.

159 Federal Supreme Court No. 134 III 166, consid. 4.2 et seq.

160 Städeli/Mary (fn 7), p. 250.

C. Other provisions

[86] It should also be noted that, in order to protect functioning competition[161], unfair competition law also sets «substantive» requirements, for example by declaring it unfair if someone disparages others and their products or business relationships by making incorrect, misleading or unnecessarily harmful statements (Art. 3 para. 1(a) UCA), or if someone causes confusion with the goods, works, services or the business operations of others (Art. 3 para. 1(d) UCA). In our opinion, however, these provisions only apply when a large language model is used in a specific application and not during training.

[87] This also applies to trade mark law, which is not dealt with separately here, but is related to unfair competition law, where it serves to protect industrial property rights. It is entirely possible for trademarks that are protected by trademark law to appear in training data. However, a trade mark right only confers on the proprietor the exclusive right to use the trade mark to identify the goods or services for which it is claimed and to dispose of it, i.e. to use the trade mark on goods or services or on business stationery (see Art. 13 of the Trademark Protection Act, **TmPA**). If training content containing another person's trade mark is used for training a language model, this does not constitute use as a trade mark and therefore cannot be prohibited under this provision.[162] However, a trademark can also enjoy copyright protection and in that case, the same applies as for works protected by copyright.[163]

D. Conclusion

[88] In terms of unfair competition law, it is advantageous if the memorisation of third-party content that could be considered the «work product» of a third party is avoided. However, even if such memorisation occurs, it can be argued that the provisions of Art. 5 UCA are not affected because only knowledge derived from these work products is used, and only if it is statistically relevant, i.e. part of general linguistic and factual knowledge that cannot be monopolised. Confidential content should also not be used for training where this has been prohibited under the relevant agreements. Content from illegal sources should also be avoided.

VIII. Excursus: Crawler bans

[89] In practice, the question of whether the increasingly common machine-readable notices on websites, according to which crawling or scraping is undesirable, must be observed (e.g. «robots.txt») is also repeatedly discussed.[164] Such notices can have relevance under copyright law, data protection law and unfair competition law:

- From a copyright law perspective, they mean that in the case of a blocking notice for crawlers, implicit consent of the website operator cannot be assumed. However, implicit

161 Federal Council's legislative commentary on the UCA 1983, p. 1039.

162 In this sense also: Berger (fn 14), SN 15.

163 See: explanations in chapter V.

164 Marmy-Brändli/Oehri (fn 14), p. 663.

consent is in any case an uncertain legal basis for crawling; the website operator would have to be the rights holder at the same time, either to assume implicit consent or to infer the absence of consent from the contradiction. In addition, as already pointed out, the scientific exception (Art. 24d CopA) applies. The purpose of exception provisions is precisely to be able to process content without the consent of the rights holder. It only requires lawful access to the works, which would also be the case if there is a blocking notice but access is not blocked. In Switzerland, such blocking notices therefore do not prevent scraping from a purely copyright perspective.

- From a data protection law perspective, a blocking notice is initially only relevant if it comes from the data subject themselves or can be attributed to them. This will not be the case in the majority of cases. If this is nevertheless the case, there is an objection to the processing of their personal data. Further processing therefore constitutes a violation of personality rights. Under Swiss law, however, such processing can be justified by an overriding interest, at least in the private sector. First and foremost, the justification of the processing for purposes not related to specific persons comes into consideration. This only fails if the personal data is memorised. In this case, it will be necessary to check whether the data subject is a public figure and therefore there is an overriding interest in the data processing. If the memorisation is sufficiently well prevented, so that only the memorisation of personal data of public figures is to be expected, then a blocking notice is irrelevant under data protection law. In the case of processing by a federal body, it is necessary to check whether the legal basis justifies the «overriding» of the objection, which will ultimately have to be checked in the context of proportionality and therefore in turn will lead to a weighing of interests in the case of proportionality in the narrower sense. The same applies *mutatis mutandis* as for private data controllers.
- This leaves the assessment under unfair competition law. Here we refer to the above explanations as to why the requirements of Art. 5(c) UCA will not generally be met. If these requirements are not met and the information is publicly accessible, we believe that the use of third-party content is not prohibited under unfair competition law, even if it is done against the express wishes of the party making the content freely accessible. This promotes competition and does not unnecessarily restrict it (which is made clear by the non-applicability of Art. 5(c) UCA). Against this background, we believe that Art. 2 UCA cannot be relevant either, which means that there is no prohibition under unfair competition law.

[90] The above also applies to crawler prohibitions in terms of use contained on websites that are not machine-readable or are not directed at robots. The reason for this is that the mere inclusion of terms of use on a website does not constitute a contract with visitors to the website and therefore does not legally bind them. For example, visitors to websites can rely on any consents expressed via such terms of use (i.e. if they permit a certain use of content), but prohibitions are irrelevant to them; in these cases, they only have to comply with the legal requirements of copyright, data protection, fair trading and other laws.

[91] The situation is different if users have to register before using the content of a website. In these cases, a contract is usually concluded in which users can undertake not to «scrape» or «crawl» the content or to use it for AI purposes (see below).

IX. Contract law

A. Breach of contract?

[92] If training content does not originate from the public internet or other public sources, it is necessary to check, irrespective of copyright law, data protection law and unfair competition law requirements, whether the party wishing to train a large language model with the relevant training data has contractually agreed not to do so or to do so only under certain conditions, such as the payment of a licence fee. No general statements can be made here, with the exception that experience has shown that most confidentiality clauses in contracts also contain a prohibition on the improper use of the other party's trade secrets. Even if memorisation can be ruled out during training or the model is not made accessible to third parties, such a prohibition on reuse prevents the data from being used for training.

[93] A confidentiality clause without purpose limitation (i.e. without a prohibition on improper use) does not prevent the training of a language model in any case if a literal and even only analogous memorisation of content (i.e. the abstracted information) can be ruled out, at least in such a way that there is no reason to assume that the output allows conclusions to be drawn about the owner of the secret (i.e. usually the business partner). A *membership inference attack*[165], which is repeatedly cited as a threat scenario for AI models, does not generally pose a problem insofar as a secret can no longer be «revealed» to someone who is already aware of it.[166]

[94] Whether the training is in breach of a contract must be determined by interpreting the contract. If the training of a language model is not clearly regulated or prohibited, it must be determined how the contract is to be understood in good faith on this point. If this leads to the conclusion that the contract contains a contractual gap, i.e. the parties have not regulated the relevant topic («May certain data be used for AI training and under what conditions? »), this contractual gap must be filled. In doing so, the dispositive statutory law on the one hand and the hypothetical will of the parties on the other must be taken into account (if it must be assumed that the parties would have supplemented the contract if they had been aware of the topic).[167] What takes precedence is subject of controversial debate.[168]

[95] Regarding the hypothetical will of the parties, we assume – subject to any provisions on purpose of use limitation – that business partners generally accept that the respective partner will gain their own expertise from the collaboration or third-party services. This is a common and accepted process, even if this know-how can be monetised later. However, it is only expected that confidentiality is maintained, i.e. that no trade secrets are disclosed to third parties and that third parties cannot draw any conclusions about the business partner.

---

165 In such an attack, an attacker tries to find out whether certain content (e.g., the data of a specific person) was used for the training, i.e., this data was a «member» of the training data. This is determined based on the answers provided by the model.

166 With further information on the Membership Inference Attack: Lisa Käde/Stephanie von Maltzan, Algorithmen, die nicht vergessen – Sicherheitslücken in Machine-Learning-Modellen und deren Bedeutung für den Schutz der Daten und der Urheberrechte, in: InTeR 2020, p. 201 et seqq., p. 203.

167 See among others: Federal Supreme Court No. 115 II 484, consid. 4b.

168 Alfred Koller, OR AT Band I, 5th ed., Bern 2023, SN 10.21: The Federal Supreme Court has repeatedly held that in the event of a contractual gap, the judge must first apply the dispositive law and only then, if this is of no further help, the hypothetical will of the parties (see e.g. Federal Supreme Court No. 115 II 484, consid. 4b). In fact, the Federal Supreme Court proceeds in the reverse order when reaching a judgment.

[96] This can also be applied to the training of large language models, as such training is comparable to the acquisition of know-how. When training large language models, this know-how is not acquired directly by the company's employees, but by the model controlled by the company. Functionally, however, both are a resource that the company can use for its other business activities. Moreover, the technical «conservation» of know-how based on business activities is nothing new. Every company that offers products and provides customer support for them will at some point build up a «knowledge base» to use the experience gained for future customer support and improvement of its products. Training an AI model is another form of preserving know-how and, in this case, with a very narrow focus on language knowledge, even if it has the advantage that it can be commercialised independently. Against this background, it can be assumed that reasonable parties would not normally prohibit each other from using the data generated during business activities to train AI models, provided that memorisation is prevented.

[97] The above may be limited where there are special commercial interests in the data material required for the know-how or training, i.e. it must be assumed that a business partner would not allow the commercial exploitation of the data material contributed by them by the other business partner without participating in the business success. This will be the case where the data material itself represents a commercial value that the company to which the data material belongs wishes to exploit.

B. Legal consequences of a breach of contract

[98] In this context, the question sometimes arises as to what the legal consequences of a breach of contractual obligations are. In principle, the following consequences are possible:

- **Termination:** If a contract is breached, at least materially, it can generally be terminated. This is of particular relevance in cases involving a licence agreement, whereby the termination of the agreement also leads to the loss of the right to use the licenced work. If a contract includes the purchase of a service, «only» the suspension of the service may be provided for instead of extraordinary termination. In this case, the contract continues to apply (including payment obligations), but the party in breach of the contract can no longer benefit from it.
- **Contractual penalty:** At least under Swiss law, the parties are free to agree a contractual penalty in the event of a breach of a specific contractual provision (Art. 160 para. 1 of the Swiss Code of Obligations, CO). Contractual penalties are not very common today. However, they are sometimes used in connection with confidentiality clauses. This is of particular relevance here, as such clauses often also provide for restrictions on the purpose of the data received from the other party to the contract. The breach of such a purpose limitation can then – depending on the wording – trigger the contractual penalty.
- **Damages:** The breach of a contract can give rise to a claim for damages (see, for example, Art. 97 CO). Damage is an involuntary reduction in assets.[169] However, it is not clear to us what this consists of in the case of unauthorised use of a business partner's data for the purpose of training a language model. In addition, the damage must be proven by the

[169] See among others: Federal Supreme Court No. 144 III 155, consid. 2.2.

business partner. Damage is conceivable if there is a memorisation and thus disclosure of the business partner's data or if the business partner becomes subject to third-party claims due to the use of the training material (because the data material originally comes from third parties). Legal costs incurred by the business partner may also constitute damages. Overall, however, the claim for damages is likely to be difficult to enforce in the present case. In certain constellations, it is still conceivable that the business partner whose contractual claims have been violated could demand the surrender of the profit that their contractual partner has made through the unauthorised use of data material. However, if a language model is not used in a commercially measurable way, this is also likely to be a difficult undertaking and the most that can be claimed in practice is saved expenses, although these would also have to be proven.

- **Other disadvantages**: In a contract, the parties can ultimately regulate the consequences of breaches of contract as they wish within the framework of the applicable law. Other consequences are also conceivable, such as the obligation to reverse the unauthorised use of the data material (e.g. by deleting the AI model). In our opinion, however, this would have to be specifically agreed.
- **Unfair competition law**: If the use of a business partner's data material for the training of language models results in a breach of contract, this may also constitute an infringement of unfair competition law. This is the case if an entrusted work product is used without authorisation (Art. 5(a) and Art. 5(b) UCA, see above in the chapter VII.B). Unfair competition law also prohibits the exploitation or disclosure of manufacturing or trade secrets that a party has found or otherwise unlawfully obtained (Art. 6 UCA). Both are also prohibited by the SPC, which criminalises both the betrayal of trade secrets and the exploitation of a third party's betrayal (Art. 162 SPC).

[99] Which of these consequences is probable or possible must be assessed on a case-by-case basis.

[100] The above statements also apply to **terms of use** of online platforms, but only insofar as the terms of use have been bindingly agreed. Under Swiss law, this is not the case when they are simply published on the online platform. In order for them to become binding contractual content, a contract must be concluded, e.g. as part of a user account registration. Anyone who can access the content of an online platform without registering or otherwise concluding a contract is not bound by any terms of use published there.[170] In this case, only – but nevertheless – the legal framework conditions for the use of such content apply. If, as is often the case, the content is protected by copyright law, it is necessary to check which of the above-mentioned principles nevertheless permit its use for training purposes. It is clear, however, that where the operator of a website objects to such use in terms of use or other notices, implicit consent cannot be assumed, at least where it can be assumed that they themselves is also the rights holder or can speak for them.[171]

[170] See further for example: Martin Eckert, Digitale Daten als Wirtschaftsgut: Besitz und Eigentum an digitalen Daten, in: SJZ 112/2016, p. 265 et seqq., p. 272.

[171] See also: chapter VIII.

C. Conclusion

[101] In practice, the most difficult hurdle for the training of large language models is the contractual regulation of access to and use of data material that is to be used for training the models. The hurdle is difficult because it has to be checked in each individual case what is permitted and because the contract often does not contain any explicit provisions and therefore has to be interpreted or even supplemented by filling in the gaps.

X. Summary

[102] Overall, the explanations show that it is basically possible to train a large language model in compliance with Swiss law using data from public sources even if the rights holders and the persons concerned have not given their express consent. However, the memorisation of training content entails the greatest risks in terms of legality, unless it can be prevented:

- If there is a verbatim or content memorisation of third-party works, certain of the commonly cited legal principles in **copyright law**, such as the scientific exception or the internal use exception, which allow the use of even public training content in the sense of an exception, may no longer apply. Here, literal memorisation (e.g. a slogan) is more problematic than mere content memorisation. However, even in these cases, the legitimacy of training a language model appears to be justifiable: In our opinion, the inclusion of such linguistic and factual knowledge in the model is not a copyright-relevant act with good reasons because the original work in the model itself either no longer occurs as a copyright-relevant work in the case of memorisation, or at least the individual character of the original work in what is present in the model as a result in the case of memorisation completely fades into the background in view of all the other information depicted in the model, and the model information has the necessary inner distance from the original works so that the corresponding rights are no longer infringed. Copyright infringement can only occur if and when an output is generated with the appropriate prompt and falls within the scope of protection of an original work. However, it is not the person who trains the model who is responsible for this, but its user; the use of a model and its output is a separate act from the training. Similar considerations apply to **unfair competition law**, at least with regard to the provisions on protection against the adoption of third-party work products.
- For **data protection law**, not only verbatim memorisation plays a role, but also analogous memorisation, in which it is not the training data itself that is memorised, but the information it contains. In this way, personal data can be incorporated into a language model. However, this memorisation does not have to be completely prevented. If public information of public figures is memorised, this will usually be justified or proportionate.

[103] Furthermore, care must be taken to ensure that no **contractual conditions** agreed in connection with the procurement or use of the training data are breached and that nothing is used resulting from a breach of a confidentiality agreement. The terms of use of online platforms (agreed as part of a registration or similar) are also deemed to be such contractual terms, even if they are publicly accessible. On the other hand, purely unilaterally expressed crawler bans are not binding for third parties, i.e. they do not have any prohibitive effects beyond the statutory provisions.

[104] Finally, we would like to point out that the considerations, interpretations and derivations set out here are largely not tested yet. For the time being, one has to live with the residual risk that the courts will come to a different conclusion. Corresponding precedents may never occur or may not occur for years. We assume that companies or entities whose business depends heavily on the protection of their content (e.g. media companies) will try to assert their interests through legal action even if the legal situation is unclear. Political initiatives can also be expected, in particular to adjust the copyright law, for example on the question of whether an «opt-out» right for the scientific exemption should be introduced in Switzerland based on EU law. Others will urge that Switzerland be strengthened as a centre of technology and research by creating even better framework conditions than those in the EU, which is – in our opinion – already the case. In our view, Swiss law allows for more possibilities regarding the development of artificial intelligence than the EU.

[105] The following figure illustrates the steps involved in testing a new source for training a large language model. Not all requirements or assumptions are covered, but those that we believe are particularly important (Figure 3, see Appendix for full size):

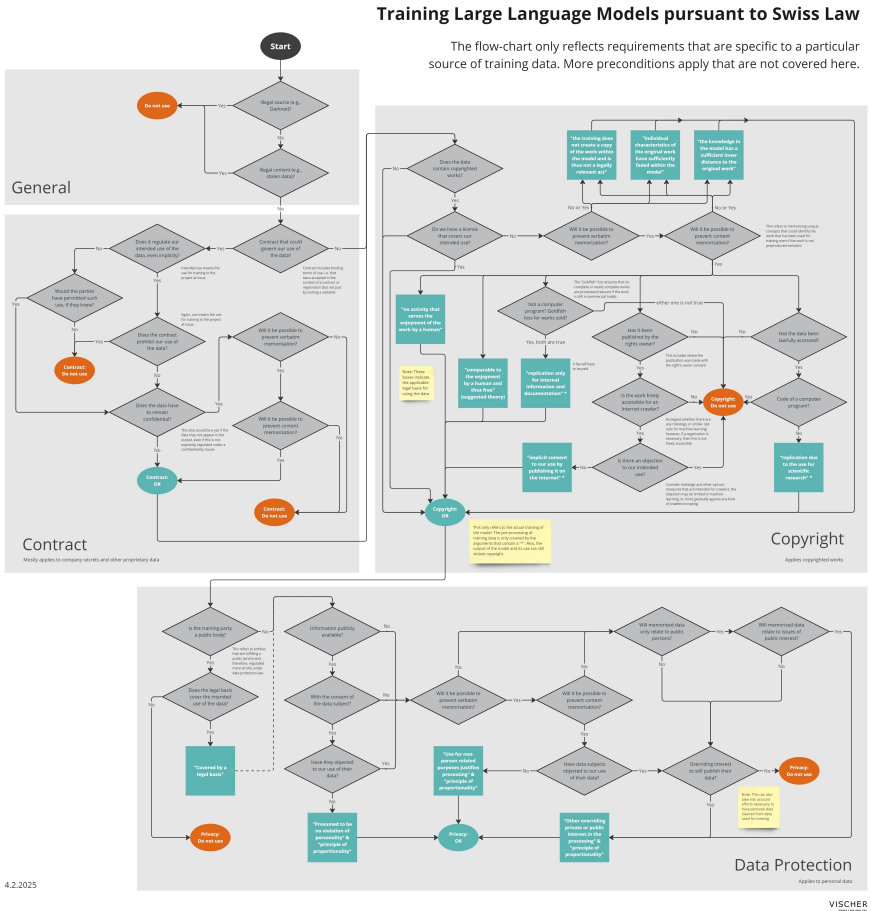

Figure 3: The testing steps involved in training a large language model under Swiss law (reduced view)

PDF Figure 3

This article is based on a memo by the authors for the ETH AI Center and represents their personal views and not necessarily those of the ETH AI Center and ETH Zurich. It was originally published in German: David Rosenthal/Livio Veraldi, Das Training von KI-Sprachmodellen mit fremden Inhalten und Daten aus rechtlicher Sicht, in: Jusletter of February 3, 2025.

David Rosenthal is a partner at VISCHER and lecturer at the University of Basel and ETH Zurich.

Livio Veraldi is a junior associate at VISCHER.

The authors would like to thank Nicole Ritter, Giulia Odermatt and Valeria Locher (all VISCHER) as well as Dr. Imanol Schlag (ETH AI Center), Prof. Martin Jaggi and Prof. Dr. Florent Thouvenin (University of Zurich) for their valuable support in the development of this article. They would also like to thank the various technical and legal experts for their valuable suggestions for this article.