|  |  |
| --- | --- |
| Modifications EDUC Moodle Course AI in Education  <https://learning.educalliance.eu/mod/hvp/view.php?id=10456&forceview=1>  Modifications made to the h5P file in the following section:    IRCT\_ AI\_M06\_24 > ETHICAL CONSEQUENCES OF AI >Which consequences of AI should we all be aware of?  Time of modification: 11/09/2024  Modified text is written with green ink | |
| Before 2024/09/11 | After 2024/09/11 |
| Openness : is AI Open ? | Openness : is AI Open ? |
| **Can so-called "open" Large Language Models be reused by anyone of us?**  When it comes to science and particularly "Open Science" which is the new paradigm of science designed to avoid frauds and increase trust in scientific results within society, the tools used should be transparent and reusable enough to let everybody replicate the experiment or the research. Do Large Language Models comply with this ideal? Mainly not because, huge resources are needed to host the datasets and the models used by AI, this kind of thing cannot be done only in somebody's garage. As Cory Doctorow puts it:  *But even the most "open" systems can't be independently replicated, due to raw computing requirements. This isn't the fault of the AI industry -- the computational intensity is a fact, not a choice (Doctorow, 2023)*  If these systems are demanding many resources it is because they are sometimes very complex to run. At a certain level, the engineers who designed them are unable to explain how they actually process information, so many are the variables taken into account by the algorithm. Is it a real problem for science?  **Yet some of these LLM can be installed on a university server...**  Nevertheless, some universities are trying to run a LLM on their servers. Their goal is to offer to all students and staff the same level of service while keeping control on the use and ecological cost of these services. A LLM itself is not enough to offer the expected level of service. Other bricks (or overalying services) are needed in order to make it work as efficiently as the products that can be found on the marketplace. Even if a LLM is scalable to a University, you will need a moderation system (among other things) to filter prompts by your students. This moderation is part of major players' business such as OpenAI (see Part 5). If your university pays for expensive subscriptions to these additionnal services, will this cost be sustainable in the long run? And in this case, can we say that universities remain independant from AI industry leaders or do they reinforce their monopoly? Is there a way out of here? | **Can so-called "open" Large Language Models be reused by any of us?**  "Open," as in "open source" or "open access," usually includes several characteristics: • It should be immediately accessible to anyone • It should be free (which doesn't mean it's necessarily gratis, but rather that we can freely use it under any conditions) • It should be transparent and auditable (think of open source code, for example)  In the next section, we will talk about transparency and reproducibility of AI Large Language Models (LLMs). At this stage, let's focus on accessibility and reusability.  On the HuggingFace website, several LLMs can be loaded without paying any fee. The filter in the left side panel allows us to select only text generation tools. Among the LLMs suggested here, most are models derived from general models which were developed by big companies. Llama was made by Facebook, the most famous GPT models are OpenAI's property, while Microsoft originally produced Phi-based models. Models whose names begin with Mistral or Mixtral were built by a French company.  Creating a completely new LLM takes too much time and resources to be accomplished by just a handful of developers, even if they are competent and enthusiastic.  Adapting one of these models for a specific series of tasks, an activity that produces some of these models derived from generic ones and is usually referred to as 'fine-tuning,' is a more achievable goal for a team of researchers or developers. However, it may still require considerable time for training and correction.  While a single person may not be able to build their own 'large language model' or may lack the resources and time to fine-tune a generic one to fit their needs, will they at least have the opportunity to load and reuse one of these LLMs on their laptop?  Obviously, this depends on the 'largeness' of the LLM, which is measured in millions or billions of parameters.  For instance, ChatGLM, a chatbot trained with data written in English and Chinese, can be used differently under different versions:  ChatGLM-130B (130 billion parameters) is far too large a model to be run from a device like a laptop and should be installed on a server (possibly run through an API). In contrast, ChatGLM-6B (6 billion parameters) can be loaded and used locally on a modern, mid-range device without requiring additional GPU power.  As a consequence, if your prompt provides much context, the 130B version will perform better than the 6B version. In this case, users must adapt their choice to their needs and desire for autonomy.  **~~Yet some of these LLM can be installed on a university server...~~**  ~~Nevertheless, some universities are trying to run a LLM on their servers. Their goal is to offer to all students and staff the same level of service while keeping control on the use and ecological cost of these services. A LLM itself is not enough to offer the expected level of service. Other bricks (or overalying services) are needed in order to make it work as efficiently as the products that can be found on the marketplace. Even if a LLM is scalable to a University, you will need a moderation system (among other things) to filter prompts by your students. This moderation is part of major players' business such as OpenAI (see Part 5). If your university pays for expensive subscriptions to these additionnal services, will this cost be sustainable in the long run? And in this case, can we say that universities remain independant from AI industry leaders or do they reinforce their monopoly? Is there a way out of here?~~ |
| TRANSPARENCY - Is AI transparent-enough (especially for science)? | TRANSPARENCY - Is AI transparent-enough (especially for science)? |
| AI are **black boxes**. But to some extent, scientists can discuss what data contributed to train the AI. If AI are trained with biased data, however complex and non interpretable the machinery, the output will have poor quality; as one says: **Garbage In, Garbage Out (GIGO)**.  Most of the time, we don't need to understand the learning algorithm in detail. Nor do we need to understand the workings of the program that the learning algorithm generates. (In so-called deep learning models, no one---including the creators of the algorithm---really understands the workings of the program that the algorithm generates.) All you have to do to spot problems is to think about the training data and the labels that are fed into the algorithm. Begin with bad data and labels, and you'll get a bad program that makes bad predictions in return. (Bergstrom, 2021)  When somebody or some corporation praises the complexity of the algorithm used by their large language model, you should consider with what data this AI was trained. If these data are of poor quality or biased, the results will be deceiving. | When speaking about data or source code, openness is closely tied to reproducibility. Reproducibility is a pillar of open and reliable science. If an experiment cannot be reproduced, its reliability can be questioned. The public's trust in Science is at stake.  Reproducibility is conditioned by transparency: every stage of the experiment should be scrutinized and kept explainable for reviewers. Is this still the case when researchers use AI tools to process data? (Let's put aside the text generation question and assume that scientists do not need AI assistants to write their text 😉)  Large model-based AI data processing tools are commonly described as unexplainable and considered black boxes. But to some extent, even if the algorithm on which the AI tool is based cannot be discussed because it's too complicated with too many parameters, it may still be possible to access and examine the kind of data used to train the model. If AI processes biased data, however complex and non-interpretable the machinery is, the output will have poor quality; as the saying goes: Garbage In, Garbage Out (GIGO).  Unfortunately, most of the time, models are made available on websites like Huggingface with poor documentation on the underlying data, and you will not usually access the dataset itself in case you would have enough time to scrutinize it, had it not been documented by someone else.  Speaking about LLMs, explainable AI tools are still closer to a myth than to reality |
| INTELLECTUAL PROPERTY - Do AI companies act as pirates? | INTELLECTUAL PROPERTY - Do AI companies act as pirates? |
| Major Artificial Intelligence Providers in the dock for copyright infringement In February 2024, The New York Times and Microsoft decided to sue OpenAI, a company whom they suspect of using their content without their consent in order to train ChatGPT (newspaper articles for the former, software stored in Github forge, for the latter) . [NVIDIA recently admitted that they had infringed copyright laws to train their LLM called NeMo](https://www.reuters.com/technology/nvidia-is-sued-by-authors-over-ai-use-copyrighted-works-2024-03-10/).  Commercial publishers invest in digital protections to prevent AI from scraping their corpuses, whether they are open access or not. As a matter of fact, Open Access licences attributed to publications suppose that when you reuse a paper, you mention the source. AI tools use many sources in so many text chunks that it is merely impossible to trace which information came from which source and credit the right sources.If things would change in the following months the question would remains: will fines of up to hundreds of millions of dollars be sufficient to compel these companies to alter their practices, especially when such significant sums of money are involved? Doubt may linger regarding this matter. How scholars and engineers can deal with it? In many cases, the question for a scientist today could be summed up like this: how can I make my publications and my data available for the public good but in the meantime prevent big AI companies from using and selling them as training content for chatbots? And what if a research team wants to develop and train its own LLM for science's sake and not for profit? Where are the corpuses that can be legally used for that purpose?  In order to help new actors to deploy new AI tools while complying with copyright rules, engineers have launched an initiative called [Common Corpus](https://huggingface.co/blog/Pclanglais/common-corpus): a large public domain dataset aimed at training AI. Maybe it's the sign that the AI industry could soon be more virtuous regarding intellectual property? | Major Artificial Intelligence Providers in the dock for copyright infringement In February 2024, The New York Times and Microsoft decided to sue OpenAI, a company whom they suspect of using their content without their consent in order to train ChatGPT (newspaper articles for the former, software stored in Github forge, for the latter) . [NVIDIA recently admitted that they had infringed copyright laws to train their LLM called NeMo](https://www.reuters.com/technology/nvidia-is-sued-by-authors-over-ai-use-copyrighted-works-2024-03-10/).  Commercial publishers invest in digital protections to prevent AI from scraping their corpuses, whether they are open access or not. As a matter of fact, Open Access licences attributed to publications suppose that when you reuse a paper, you mention the source. AI tools use many sources in so many text chunks that it is merely impossible to trace which information came from which source and credit the right sources.  Large companies like OpenAI frequently violate copyright laws on a daily basis. Tribunals and society in general seem to act far more harshly against those who infringe copyright by letting scientific results return from commercial providers to the community that produced them, than against companies that violate copyright laws by exploiting all available texts to train their AI models with no regard for licenses  If things would change in the following months the question would remain: will fines of up to hundreds of millions of dollars be sufficient to compel these companies to alter their practices, especially when such significant sums of money are involved? Doubt may linger regarding this matter. How scholars and engineers can deal with it? In many cases, the question for a scientist today could be summed up like this: how can I make my publications and my data available for the public good but in the meantime prevent big AI companies from using and selling them as training content for chatbots? And what if a research team wants to develop and train its own LLM for science's sake and not for profit? Where are the corpuses that can be legally used for that purpose?  In order to help new actors to deploy new AI tools while complying with copyright rules, engineers have launched an initiative called [Common Corpus](https://huggingface.co/blog/Pclanglais/common-corpus): a large public domain dataset aimed at training AI. Maybe it's the sign that the AI industry could soon be more virtuous regarding intellectual property? |
| COGNITIVE SKILLS - Should Universities ban AI use by strudents? | COGNITIVE SKILLS - Should Universities ban AI use by strudents? |
| Most of the time universities don't really prohibit the use of artificial intelligence by their students, but they impose rules that encourage students to be transparent in their use. There's nothing more irritating for a teacher than correcting ChatGPT while they are supposed to be correcting a student's original production.  The same problems apply to researchers: artificial intelligences make it possible to dispatch tedious tasks and save time, but no one is immune to so-called “hallucinations” (a marketing term coined by generative AI providers to qualify the distance between probabilistic output of the tool and what we know of ground reality) If the tool is not properly adapted to the task, even the slightest inexplicable error of these artificial intelligences can have enormous consequences.  The link below sets out contradictory arguments. It's up to you to vote for these arguments. Compare your votes with those of other students who have done this activity. | Most of the time universities don't really prohibit the use of artificial intelligence by their students, but they impose rules that encourage students to be transparent in their use. There's nothing more irritating for a teacher than correcting ChatGPT while they are supposed to be correcting a student's original production.  The same problems apply to researchers: Artificial intelligence systems make it possible to dispatch tedious tasks and save time, but no one is immune to their so-called 'hallucinations' (a marketing term coined by generative AI providers to describe the discrepancy between the probabilistic output of the tool and what we know of ground reality). If the tool is not properly adapted to the task, even the slightest inexplicable error of these artificial intelligences can have enormous consequences.  That said, using Generative Artificial Intelligence tools can be highly beneficial with some control and experience:  You can use them to practice (prompt them to ask you questions about a specific text and correct your answers) and to generate ideas when starting with a general question (ideation). The more experienced you are with the subject matter, the more benefit you will draw from this activity. You cannot properly use text generated by an AI if you are not able to critically evaluate and correct the output.  The link below sets out contradictory arguments. It's up to you to vote for these arguments. Compare your votes with those of other students who have done this activity. |
| SOCIAL IMPACT - Do AI corporations run sweatshops to moderate the content of their products? | SOCIAL IMPACT - Do AI corporations run sweatshops to moderate the content of their products? |
| The stock market valuation of artificial intelligence seems to know no bounds. But, as with other extractive activities, these immense profits depend on the exploitation of cheap foreign labor. In this case, the Times revealed in January 2023 that OpenAI had paid Kenyan moderators less than $2 an hour to make the content used by ChatGPT less toxic.  The click workers and other mechanical turks who are so necessary to AI training and often have a high level of education frequently earn less than the local wage for an equivalent number of hours spent working.  AI is also arriving everywhere in the workplace, and employees generally have no opportunity to oppose it. A time-saving tool, Artificial Intelligence can also sometimes turn into a surveillance tool for the employee, summoned to adapt to the tool when it should be the tool that adapts to the worker.  Digital Labor is not only mobilized by developers to filter datasets of any kind, they are also recruited to classify and label the millions of items which these datasets consist of. In the process, prejudices and stereotypes belonging to this invisible army of workers, hired through the Amazon Mechanical Turks platform, are transferred into the classification. This has been the case for years until the ImageNet dataset was made available online.  "Over several years of development, ImageNet grew enormous: the development team scraped a collection of many millions of images from the internet and briefly became the world's largest academic user of Amazon’s Mechanical Turk, using an army of piecemeal workers to sort an average of 50 images per minute into thousands of categories"  (Kate Crawford and Trevor Paglen, https://excavating.ai) | The stock market valuation of artificial intelligence seems to know no bounds. But, as with other extractive activities, these immense profits depend on the exploitation of cheap foreign labor. In this case, the Times revealed in January 2023 that OpenAI had paid Kenyan moderators less than $2 an hour to make the content used by ChatGPT less toxic.  The click workers and other mechanical turks who are so necessary to AI training and often have a high level of education frequently earn less than the local wage for an equivalent number of hours spent working.  AI is also arriving everywhere in the workplace, and employees generally have no opportunity to oppose it. A time-saving tool, Artificial Intelligence can also sometimes turn into a surveillance tool for the employee, summoned to adapt to the tool when it should be the tool that adapts to the worker.  Digital Labor is not only mobilized by developers to filter datasets of any kind, they are also recruited to classify and label the millions of items which these datasets consist of. In the process, prejudices and stereotypes belonging to this invisible army of workers, hired through the Amazon Mechanical Turks platform, are transferred into the classification. This has been the case for years until the ImageNet dataset was made available online.  "Over several years of development, ImageNet grew enormous: the development team scraped a collection of many millions of images from the internet and briefly became the world's largest academic user of Amazon’s Mechanical Turk, using an army of piecemeal workers to sort an average of 50 images per minute into thousands of categories"  (Kate Crawford and Trevor Paglen, https://excavating.ai) |
| CLASSIFICATION is political | CLASSIFICATION is political |
| AI training data sets must be categorized by digital workers in order to become reusable by text-to-image generation tools. The biases of these "click-workers" are reflected in the output of these tools. Thus, the use of tools to generate images convey a large risk to strenghten stereotypes.  Imagenet, a large dataset of images, created in 2009 and withdrawn from the net in 2019 was intensively used to train many AI tools. These databases orriginally consisted of +21000 categories ranked under [17 top-level categories](https://web.archive.org/web/20200907212153/http:/image-net.org/about-stats.php) as different as "persons", "fungus", "vehicles" or "furniture"... the majority of these categories are not at all controversial (the distinction between banana and apple cannot cause any harm to anybody). But when it comes to classifying people, things usually get weird and very sensitive. The database used to contain up to 3000 sub-categories, under the top-level category "people". Among these categories, there was no place between "male" or "female" for anything else (the category transgender was absent). The category "hermaphrodite" was present but in another branch, "sensualist" (a term that conveys moral defect), and applied to a portrait of the actress Sigourney Weaver.  In Imagenet, this portrait was labelled "loser": | AI training data sets must be categorized by digital workers in order to become reusable by text-to-image generation tools. The biases of these "click-workers" are reflected in the output of these tools. Thus, the use of tools to generate images convey a large risk to strenghten stereotypes.  Imagenet, a large dataset of images, created in 2009 and withdrawn from the net in 2019 was intensively used to train many AI tools. These databases orriginally consisted of +21000 categories ranked under [17 top-level categories](https://web.archive.org/web/20200907212153/http:/image-net.org/about-stats.php) as different as "persons", "fungus", "vehicles" or "furniture"... the majority of these categories are not at all controversial (the distinction between banana and apple cannot cause any harm to anybody). But when it comes to classifying people, things usually get weird and very sensitive. The database used to contain up to 3000 sub-categories, under the top-level category "people". Among these categories, there was no place between "male" or "female" for anything else (the category transgender was absent). The category "hermaphrodite" was present but in another branch, "sensualist" (a term that conveys moral defect), and applied to a portrait of the actress Sigourney Weaver.  In Imagenet, this portrait was labelled "loser": |
| ENVIRONMENTAL ISSUES - Is AI the new pipeline (and data the new oil)? | ENVIRONMENTAL ISSUES - Is AI the new pipeline (and data the new oil)? |
| To say that artificial intelligence is an extractivist activity means at least two things. First, it proceeds similarly to a mining activity, extracting digital traces of our identities and behaviors from our digital lives. Secondly, it is inseparable from mining in the most physical sense of the term. To continue developing, the AI-based economy needs to extract more and more rare earth elements from the ground, destroying landscapes in the process, and discarding anything that cannot be recycled in the form of sludge.  Kate Crawford in her book "Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence" published in 2021 explores the two sides of this extractivism on which the AI industry is based.  What's more, the imposing servers that enable AI to function need water to be cooled (169 liter/second for a Google Datacenter (Sebastian Lehuédé, 2024).  As a consequence, there may be serious water shortages for the surrounding populations, as is the case in Chile where there is no public policy on this essential resource.  Finally, training artificial intelligence and then using it on a day-to-day basis consumes a considerable amount of energy and causes higher greenhouse gas emissions than older technologies that remain in service.  All these factors should encourage us to use artificial intelligence sparingly. Can we really afford to use it all the time?  Can we really use chatbots everytime we feel the need to use them, or should we ask ourselves if our own intelligence, which is far less energy-consuming and much more sophisticated than ChatGPT, is enough to perform simple tasks, such as writing a text. Writing something for ourselves and for others after all is the best way to test if we have really understood what we are talking about.  The link below leads to an activity inspired by the publication of a 2023 preprint. The study tries to measure the cost per inference of popular artificial intelligences of different types. Cost is measured in number of loads for an average smartphone (Luccioni, Yernite et al., 2023).   * Inference = a prompt/question given to a chatbox * Load = a full battery charge   Try to associate each type of AI with its actual energy cost. | To say that artificial intelligence is an extractivist activity means at least two things. First, it proceeds similarly to a mining activity, extracting digital traces of our identities and behaviors from our digital lives. Secondly, it is inseparable from mining in the most physical sense of the term. To continue developing, the AI-based economy needs to extract more and more rare earth elements from the ground, destroying landscapes in the process, and discarding anything that cannot be recycled in the form of sludge.  Kate Crawford in her book "Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence" published in 2021 explores the two sides of this extractivism on which the AI industry is based.  What's more, the imposing servers that enable AI to function need water to be cooled (169 liter/second for a Google Datacenter (Sebastian Lehuédé, 2024).  As a consequence, there may be serious water shortages for the surrounding populations, as is the case in Chile where there is no public policy on this essential resource.  Finally, training artificial intelligence and then using it on a day-to-day basis consumes a considerable amount of energy and causes higher greenhouse gas emissions than older technologies that remain in service.  All these factors should encourage us to use artificial intelligence sparingly. Can we really afford to use it all the time?  Since the beginning of this century, Google has been increasing its emissions faster than the average industrial stakeholder. This has led us to consider whether, instead of typing "New York Times" in Google Search, we should rather rely on our memory and type "nyt" in the address bar to let the browser complete the query. Today, we have to ask ourselves if we really need to use an AI chatbot to get the answer we need, as it's a far less ecological means than typing in a search bar. (It's worth noting that the 2024 Google Search version is now AI-powered, and it's becoming difficult to use an AI-free Google Search version.)  Can we really use chatbots everytime we feel the need to use them, or should we ask ourselves if our own intelligence, which is far less energy-consuming and much more sophisticated than ChatGPT, is enough to perform simple tasks, such as writing a text. ~~Writing something for ourselves and for others after all is the best way to test if we have really understood what we are talking about.~~  As Epicurus said, we need to 'cultivate our garden,' which in a digital world means recognizing our true digital needs (not those imposed by industry), choosing less resource-consuming tools that meet those needs, and being more conscious of the technologies that genuinely enhance our digital skills. We need fewer chatbots and more digital literacy.  The link below leads to an activity inspired by the publication of a 2023 preprint. The study tries to measure the cost per inference of popular artificial intelligences of different types. Cost is measured in number of loads for an average smartphone (Luccioni, Yernite et al., 2023).   * Inference = a prompt/question given to a chatbox * Load = a full battery charge   Try to associate each type of AI with its actual energy cost. |