Reflection Assignment
AI and Your Profession

Maja Günther
Student Nr. 506848

October 27, 2025

Contents

1 Week 2 - Course introduction, Modelling Intelligence, and AI subfields **3**
1.1 What parts of society can be represented in a NetLogo model, and what important complexities are inevitably left out? . . . 3
1.2 If a NetLogo model simulates wealth distribution or social networks, how might the assumptions reinforce stereotypes or overlook systemic inequalities? . . . 3
1.3 What are your key takeaways? . . . 5
1.4 Will you be applying these concepts in your future profession? If so, how? . . . 5

2 Week 4.1 - State of the Art and Introduction to LLMs **5**
2.1 Research the use of AI in the UN Sustainable Development Goals. . . . 5
2.1.1 Which of the Sustainable Development Goals are of biggest importance in your future profession? . . . 6
2.1.2 Based on what you've read, do you believe that AI can help obtain these goals? . . . 7
2.2 Based on what you've learned about Large Language Models such as ChatGPT, Copilot and Gemini, where on this spectrum would you say such models fall? . . . 7
2.3 Do you agree with this reasoning? Why (not)? . . . 8
2.4 What are your key takeaways? . . . 8
2.5 Will you be applying these concepts in your future profession? If so, how? . . . 8

3 Week 4.2 - Ethics and Trustworthy AI #1 **8**
3.1 Which ethical issues may arise in your professional field due to the application of AI or data science? Name at least 2 such issues to consider. . . . 8
3.2 In the lecture, we discussed a framework for trustworthy AI proposed by IBM, in particular the five pillars that they consider. . . . 9
3.2.1 To what extent do you expect to encounter these five pillars in your professional field? Do any of them stand out or are all five equally important? . . . 9
3.2.2 Link each of the ethical issues you investigated for question 1, to one or more of these pillars. . . . 9
3.3 What are your key takeaways? . . . 9
3.4 Will you be applying these concepts in your future profession? If so, how? . . . 9

4 Week 5.1 - Guest lecture by Isabelle Tilleman **10**
4.1 Have a look at the latest version of the SRIR roadmap that TNO has worked on with the EU. . . . 10
4.1.1 Pick out at least one term in the roadmap that you did not know before. Research what it means. Is this a relevant term for your field of work or not at all? Explain your answer. . . . 10
4.1.2 The TAILOR project identifies four topic sectors. To what extent does each topic sectors have a relation to your field of work? Explain your answers. . . . 10
4.2 What is your view on this? Should the EU prioritize accelerating innovation, even if it means loosening its emphasis on AI governance and ethics? . . . 11
4.3 What are your key takeaways? . . . 11
4.4 Will you be applying these concepts in your future profession? If so, how? . . . 11

5 Week 5.2 - Trustworthy AI #2 **11**
5.1 Assignment on ALTAI. Select a specific data-driven or AI-based application in your professional field. . . . 11
5.2 Then, make an assessment of the ethical aspects concerning the application. Use the assessment list for trustworthy AI (ALTAI) as discussed in the lectures. . . . 12
5.3 Of course, not every item on the ALTAI is of equal relevance. Show that you can identify the most relevant ethical items for your selected application. . . . 12
5.4 What are your key takeaways? . . . 13
5.5 Will you be applying these concepts in your future profession? If so, how? . . . 13

6 Week 6.2 - Guest lecture on legal aspects of AI by Trix Mulder **13**
6.1 The AI act recognizes a number of categories of sensitivity. Argue which category your AI application falls into. . . . 13
6.2 Are there any issues around copyright or ownership of data in your project assignment? 14
6.3 What is your personal opinion on the AI act? . . . 14
6.3.1 Does the law go far enough, in your opinion? . . . 14
6.3.2 Do you think the law will be effective? . . . 15
6.4 What are your key takeaways? . . . 15
6.5 Will you be applying these concepts in your future profession? If so, how? . . . 15

7 Week 7.1 - Guest lecture on prompt engineering by Jos Punter **16**
7.1 Which prompt engineering techniques did you find most useful and why? . . . 16
7.2 Give an example of a prompt that you have improved by applying prompt engineering techniques. Show the results before and after and explain what has improved in the results. . . . 16
7.3 Read this blog on green prompting. What do you think of this concept? . . . 17
7.4 What are your key takeaways? . . . 17
7.5 Will you be applying these concepts in your future profession? If so, how? . . . 17

8 Week 7.2 - Guest lecture on AI Standardization by Willy Tadema & Theresa Marschall **18**
8.1 Do you think this timeline is realistic, based on what you have seen so far in practice? For example, in the company you're working with for your group project, or a different company in your field. . . . 18
8.2 What are your key takeaways? . . . 18
8.3 Will you be applying these concepts in your future profession? If so, how? . . . 18

1 Week 2 - Course introduction, Modelling Intelligence, and AI subfields

Follow the NetLogo tutorial from the class, and play around with some different models of your choosing. Particularly, have a look at some societal models.

I think it's amazing how many models NetLogo provides for free use! I wasn't able to find a number online and figured my time counting is probably better spent writing, but I am totally amazed! I tried out quite a few models and I especially liked the Paths one from the Social Science section. I also worked on this while on a table with psychology students and they were so curious.

I feel like some of the use cases are quite abstract to understand, at least if you don't have someone explaining the variables and I don't really like, how you cannot see the simulation while looking at the description, but obviously that's a me problem.

1.1 What parts of society can be represented in a NetLogo model, and what important complexities are inevitably left out?

NetLogo models can represent variables that we are already observing and collecting data about, like traffic, infection rates, etc. Factors that cannot be represented are mostly also factors, that we cannot predict in real life, or that are at least hard to incorporate into a model. To represent human emotion or cultural differences in the societal models would require an endless amount of factors, which cannot be expected. Furthermore the predictions can always be thrown overboard by, for example, extreme weather events, an economic collapse, maybe even just one of your friends being in bad mood, if you tried to predict the outcome of your party with the NetLogo Party model.

This made me think of the Pentagon Pizza theory recently that gained attention again, stating broadly that, the more pizza gets delivered to the Pentagon, the more likely it is, for a big, often war-related event to happen soon. (Mouriquand, 2025)
How fun would it be if a NetLogo model would include a scrollbar for the amount of pizzas delivered to the Pentagon, so you can REALLY predict your use case.

1.2 If a NetLogo model simulates wealth distribution or social networks, how might the assumptions reinforce stereotypes or overlook systemic inequalities?

I have a lot to say to this question but I am struggling A LOT to put it into a coherent entry.

Firstly, after looking into the model info of the Wealth Distribution model, I figured, that it is quite hard to understand all the customizable values, their impact not just on the whole outcome but also on each other. So if you are not an expert, working with this model may predict complete nonsense. I am very impressed by the whole website and the amount of models they provide, but if I was supposed to work with these models for uni or do a whole submission with it, I would expect a deep explanation from my lecturer.

Then, of course the model's assumptions will reinforce stereotypes or overlook systematic inequalities, if they are used by people/companies who aren't sensitized on said issues. If the provided data is already biased, the bias will persist. If the evaluating person isn't sensitized on said said issues, they probably won't look further into it in their report.
As NetLogo appears to be an US company, I think we can assume, that their research and data also widely focuses on american and western culture and behaviour. I have a map in mind, that shows which countries most data/statistics come(s) from and puts it into context with the population share. I wasn't able to find it, so I compromised on Figure 1. While looking for the original publication/higher resolution, I then stumbled upon an article with not only this figure in higher resolution but also the graph I was thinking of, so I will talk about both.

Figure 1 is titled "Which Countries Are Leading the Data Economy?". It shows the data accessibility on the X-axis and and the broadband consumption per internet user in terabytes on the Y-axis. Countries that rank higher in data accessibility and broadband consumption per user are considered "winners", with the US scoring the highest values on both axes.

While I was able to find Figure 2, there was no proper source provided and the legend is not included in the graphic, but instead in the caption. I linked the article I got it from, since there is nothing else to link. This is the Top 10:

1. United States
2. United Kingdom
3. China
4. Switzerland
5. South Korea
6. France
7. Canada
8. Sweden
9. Australia
10. Czech Republic

The graph basically just shows, that certain countries collect and share (significantly) more data than others. If you are from one of the gray countries, you probably don't even have to look at NetLogo data and it's inaccuracies, but simply most statistical findings made with public available data, to figure out, that your culture and beliefs aren't properly represented here.

Another example, that wouldn't have required a long scavenger hunt for a map of unsupported data would've simply been the Party example again. The model doesn't allow me to look at factors like sexual orientation, cultural background, hobbies, other similarities of my guests. These are factors that violently affect group building.

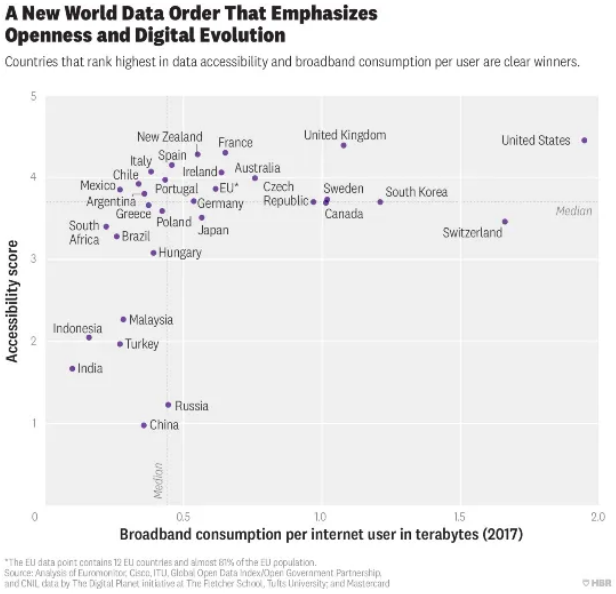

Figure 1: (Bhaskar Chakravorti & Chaturvedi, 2019)

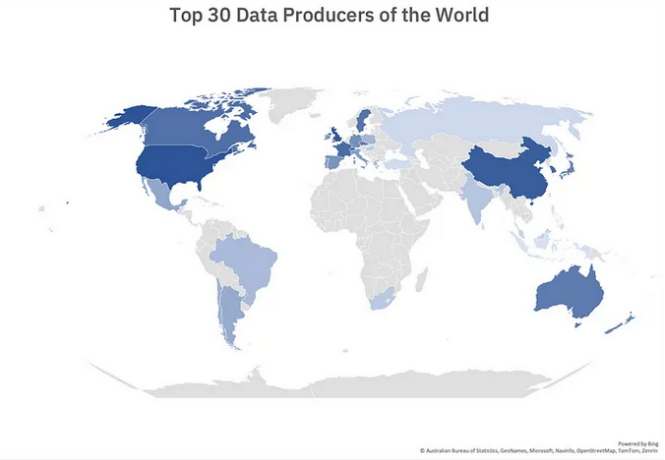

Figure 2: (Bhattacharyya, 2023)

1.3 What are your key takeaways?

My key takeaway from the first lecture was, that AI is just such a wide range topic and there is always so much more to talk about and to go into detail with. I have had about three introduction lectures to AI at different universities so far and they all differ from depth and topics.

There are four types of AI by awareness:

- Reactive machine
- Limited memory
- Theory of mind
- Self awareness

And the subfields we talked about were:

- Robotics/Autonomous Systems
- Computer Vision
- Speech/Audio processing
- Natural Language Processing
- Expert Systems
- Planning and Organization
- Machine Learning

During our exercise to give an example of each category in a different field, we had trouble finding examples for Energy (or rather distinguishing the difference between Energy and Sustainability).

Also my actual most present key takeaway was the 8 queens problem, because I love playing the Queens LinkedIn Mini Game, but I haven't made the chess connection until that very day. I think about it daily since (whenever I have my mini games break) and I also love telling people about it, now that I know.

1.4 Will you be applying these concepts in your future profession? If so, how?

My future profession – at least directly after my bachelors – will most definitely be as a Software Engineer of the Identity Management department at the data centre of the Free University of Berlin. Since university employees are part of the public sector, there are quite strict rules about AI usage: No AI for any confidential data.

I think it is very hard to introduce such a big and oncoming topic into an environment like my workplace – a small datacentre with around 60 employees, most of them working there since 10+ years. And also introducing AI in such a privacy and bureaucracy loving country as Germany will most likely pose as a problem. I do know that my colleagues in the Scientific Computing department have a big interest in AI, but due to the Cooperative Study program plan, it's kind of hard to break into new departments when you haven't worked for them before.

But for Computer Science as a field of work in general, AI is of course gaining more and more importance on a daily basis. I doubt there will be a way to dodge AI in that field. Even if you aren't actually working with it, everyone is always updated on the newest technologies and inventions.

2 Week 4.1 - State of the Art and Introduction to LLMs

2.1 Research the use of AI in the UN Sustainable Development Goals.

While AI is driving innovation and redefining solutions to complex problems, realizing those benefits requires intentional design, responsible governance and inclusive participation. While it is nice to profit off the constantly evolving technology, the prospect of the SDGs will always be *meeting the needs of the present without compromising the needs of future generations*.

Benefits, for example,	Risks, for example,
- improved climate foresting	- surveillance concerns
- fraud detection	- absence of clear ethical guidelines
- strengthened transparency	- training with biased data
- creating jobs	- contribution to carbon emissions

The UN Global Compact encourages companies to align their AI use with the SDGs. It also challenges them to maximize their contributions to the 2030 Agenda with AI. (UN Global Compact, 2025)

2.1.1 Which of the Sustainable Development Goals are of biggest importance in your future profession?

Of course all the Sustainable Development Goals are important for everyone in all aspects - otherwise they wouldn't have been declared a goal.

I think working at a university, especially Number 4 - Quality Education is important in my future profession. The Free University Berlin is always working on providing affordable but high quality education to everyone. University fees in Germany are also not nearly as high as in the Netherlands. I think the current semester fee is around 300€, so 600€ a year, which does include a public transport ticket, covering all trains, metros, trams and busses in the whole country, except for the highspeed trains. If you feel you don't need this ticket you can apply for an exemption. While the General Students Committee usually makes the agreements, they work closely with the universities Identity Management, so all data can be transmitted correctly and safely.
Working in Identity Management also lays weight on 10 - Reduced Inequalities. The department does not only work with the simple transmission of student data, but also equality concerns. Students requests about changing display name and gender entry go through the department, gender neutral language in any communication is handled here and a recent problem we encountered right before I left, is the handling of certain characters. While I already tend to struggle with the ü in my name here in the Netherlands (try explaining to the post office staff, that the name on the package is probably not Günther but instead G%UUMLnther), it is important to make sure, that there will be no such issues happening in the university handling.

I also like to think, that the Free University Berlin does concern itself with with Number 16 - Peace, Justice and strong Institutions, as its motto is: *Veritas, Iustitia, Libertas*, which translates to Truth, Justice, Liberty.

For the Computer Science field overall, I think 12 - Responsible Consumption and Production are a nice goal for an accountable industry. Using resources efficiently and minimizing waste and pollution throughout a product's lifecycle should be every industries goal. Of course 8 - Decent Work and Economic Growth is nice for me as an undergraduate student, but I do feel like that's more for the entrepreneurial part of the program, who would all like to be self-employed working from Bali soon. As I am graduating with three years of work experience, I would like to think I'll be fine. I know that I will never manage my ADHD well enough to be my own boss, and I am very well adaptable to different work environments.

Lastly, in a more broad way, I think as a woman in a male dominated field, Number 5 - Gender Equality is an issue that will persist in my future profession, already right now, but also if I end up leaving the Free University datacentre. Computer Science persists to be an unattractive field for women, in my experience rather due to the people than the actual study. My workplace employs quite a few women, but significantly more for administrative work than IT, the AIP program has four women, out of twenty students, my CP program in Berlin counts six women out of thirty students.
I have been told by my classmates, that its of biological nature, that women simply are less interested in STEM. In my first weeks at work, I had multiple colleagues tell me I must be wrong, since this is the datacentre. My classmate has been comforted by a work supervisor, that I only got a better grade, because I'm a girl and that he shouldn't be worried.
Speaking up in such situations is obviously necessary and that will persist for the rest of my worklife I fear.

2.1.2 Based on what you've read, do you believe that AI can help obtain these goals?

I think that AI can definitely help anyone interested obtaining the SDGs. I just fear that there are much more people using AI for more profit driven goals, who couldn't care less about the SDGs.
If used efficiently, AI can definitely help providing Quality Education, helping users with customized learning plans, detecting individual weaknesses and strengths and working with those. Most AI applications and also regulations we are currently using will probably not do much for reducing inequalities, to say it frankly. Bias in training data is almost always existent and if the people who are supposed to overlook the system aren't unbiased themselves, they probably won't notice. (My later mentioned law student friend once took a class, that taught them to beware, that most AI systems will *always* target the black man.) But that's something we can work on, hopefully!
For my portfolio I read an essay titled "Why AI will save the world" (Andreessen, 2023), from a man who is very convinced, that AI systems should be used in war, since they will finally help officials make smarter strategies. I would like to cite Alan Jacobs note to said paragraph here, in context of Number 16 - Peace, Justice and strong Institutions:

> `M.A. does not seem to understand the concept of war. Nations that go to war are typically not interested in reducing bloodshed; they don't end up killing a lot of people because they have a shortage of information. They *want* to kill a lot of people. If Putin could kill every Urkainian he probably would. Would advanced A.I. help him to do that? You bet it would.`

Goal 12 - Responsible Consumption and Production could be reached with the help of AI, but probably faster without actually. AI chatbots do everything but encourage consumers to responsible behavior. Also that one will probably clash with Goal 8 - Decent Work and Economic Growth. AI can definitely help with economic growth, and a common argument (also in the horrible essay) is, that while AI will replace jobs, it will always create new ones, the new ones can probably not be considered "decent". I don't think AI will provide any significant breakthroughs in gender equality.

In the second week, we identified four different "stages" of awareness in types of AI. Read the article attached to the slides on these stages.

2.2 Based on what you've learned about Large Language Models such as ChatGPT, Copilot and Gemini, where on this spectrum would you say such models fall?

The article defines four stages based on Functionalities:

1. Reactive Machines: responds to current input only, no data storage
2. Limited Memory: uses past data and experiences
3. Theory of Mind: understands emotions and beliefs
4. Self-aware AI: is conscious and self-reflective

Large Language Models like ChatGPT, Copilot and Gemini definitely belong to the Limited Memory stage. They can use context from past data and previous conversations to generate better fitting results. They do not understand emotions, intentions or possess consciousness, but many are trained to react to obvious emotion indicators (writing in caps, using swear words, expressing discontent).
Based on Capabilities it differentiates in three types:

1. Narrow AI (Weak AI)
2. General AI (Strong AI)
3. Superintelligent AI

Our examples are considered Narrow (Weak) AI, because they are powerful in language tasks but not capable of human-like reasoning or self-awareness. LLMs are advanced forms of AI, but they are still far from "Theory of Mind" or "self-aware" intelligence.

Read this Medium blog/article from an author who believes that calling the "hallucinations" we discussed in class, is a mistake because it humanizes an LLM.

2.3 Do you agree with this reasoning? Why (not)?

I do agree with him, that from a developer view, hallucination is the wrong word and does make it sound like a binary malfunction rather than a mathematical process. But in my understanding, hallucination is a word much rather used to explain to consumers, not developers, why AI may not always be their answer. If you're trying to convince a middle school student to write an essay without ChatGPT, saying "[it's] executing a stochastic sampling process based on learned probability distributions" won't help much. "It's gonna make up stuff and you wouldn't even know" sounds much more convincing and appropriate.
So I, as a third year CompSci student, agree with his reasoning, but I don't think it's suitable for educating about AI risks, as it assumes too much prior knowledge.

2.4 What are your key takeaways?

Types of AI Hallucinations are:

- sentence contradiction
- Prompt contradiction
- Factual contradiction
- Nonsensical/irrelevant contradiction

And common causes of said Hallucinations can be:

- Poor data quality
- Text generation method
- Input context confused the model

2.5 Will you be applying these concepts in your future profession? If so, how?

It is so important to check AI for hallucinations when using it in the workplace.
Using AI for tasks like writing emails or making presentations, when you don't provide them with sufficient context (maybe confidential work data), will force it to hallucinate. So if the data is left out "on accident", that may lead to an uncomfortable reply from your boss. As I work in a very chill environment, I am writing my emails myself, but I do understand when young people, working their first professional job, let an AI "proofread" their emails. I aspire to never work in an environment, that will carry me to do that.

3 Week 4.2 - Ethics and Trustworthy AI #1

3.1 Which ethical issues may arise in your professional field due to the application of AI or data science? Name at least 2 such issues to consider.

I think the biggest AI risk for CompSci graduates is simply the fact, that many companies like to think they can replace employees with AI, so the decrease of available job positions.
Also the increasing amount of vibe coding leads people to mindlessly paste confidential data and company secrets into chatbots. We have had to renew passwords not only because of newly employed students, but also senior employees. Thankfully my workplace uses a keystore!
Also using AI for complex problems instead of bothering a classmate/supervisor/teacher often leads to less understanding of the used code. If you use ChatGPT code you don't understand once, because it solves your problem, you'll most likely not understand if it's also the cause of the next problem you encounter.

3.2 In the lecture, we discussed a framework for trustworthy AI proposed by IBM, in particular the five pillars that they consider.

The five pillars are: Explainability, Fairness, Robustness, Transparency and Privacy.

3.2.1 To what extent do you expect to encounter these five pillars in your professional field? Do any of them stand out or are all five equally important?

I think all of the five pillars matter equally for the CompSci field. I like to think that Privacy and Robustness are such important principles, that applications who cannot live up to the standard won't even make it to the evaluation of the other factors, but data breaches happen more often than you'd think, so maybe wrong perception on my part.

3.2.2 Link each of the ethical issues you investigated for question 1, to one or more of these pillars.

I don't think the job problem is linkable to one of the pillars.
The data/password problem is definitely linked to privacy: many companies let their employees use the companies own AI, because they have gone through the process to ensure, that no confidential information will be published. Internal AI keeps data processing within the company systems. This provides full control over security, access policies, and privacy.
I think the reliance on code you cannot understand is most likely linked to Explainability. If the user is not able to understand what is happening, how can they then apply it? (I do blame user behavior more than the AI here)

3.3 What are your key takeaways?

My **key** takeaway from that lesson was the existence of the "Can I wear shorts today" Website, that Ruben introduced us to.
I really liked the discussion questions form that lecture. So for example, the argument that the standard for AI systems shouldn't be no bias, since the humans whose work they're supposed to replace weren't unbiased either makes absolutely no sense, BECAUSE that's exactly the opportunity of AI we should be embracing!
I think it was a shame that that lecture was at 8:30, because this was a really interesting! (I just drifted off researching if the ITBM major includes ethics courses, which proved to be harder than imagined, considering I don't speak dutch)
I also wrote down the recommended book for my learning plan, as it sounded very interesting. Have not read it at this point in time.

3.4 Will you be applying these concepts in your future profession? If so, how?

I think so! I think the principles and requirements we learned about are quite logical. It's nice to keep them in mind, not only when actually implementing stuff, but also just when using AI apps or reading advertisements.

4 Week 5.1 - Guest lecture by Isabelle Tilleman

4.1 Have a look at the latest version of the SRIR roadmap that TNO has worked on with the EU.

4.1.1 Pick out at least one term in the roadmap that you did not know before. Research what it means. Is this a relevant term for your field of work or not at all? Explain your answer.

The term I picked is: **Federated Learning**.
Federated learning is a decentralized and privacy-friendly form of machine learning. There is no need for a central database to hold all of the sensitive data. Instead of bringing the data to the machine learning model, Federated Learning brigs the machine learning model to the data. In this way, the training of the models is broken down into sub-calculations that are performed locally at an organization. After carrying out the calculations, only the anonymized results are shared with the organizations conducting the research, not the privacy-sensitive data itself. It solves two major problems of data analysis: improved qualitative analyses for society and safeguarding of citizens' right in relation to privacy. At the current date, analyzing large amounts of data is easier than ever before. Computing power is increasing and algorithms are becoming ever more advanced.
It allows companies to harness data without violating privacy. Federated learning leads to better predictions and better models. (Molhoek, 2025)
I think this is a relevant term for my field of work, because it allows collaborative machine learning without directly sharing data, addressing privacy and security challenges. It also drives innovation in distributed systems, data management, and AI model optimization—key areas within modern Computer Science.

4.1.2 The TAILOR project identifies four topic sectors. To what extent does each topic sectors have a relation to your field of work? Explain your answers.

The four topic sectors are:

1. Trustworthy AI (TAI)
2. Learning, Optimization, and Reasoning (LOR)
3. Ethical, Legal, and Societal Aspects (ELSA)
4. Data and Infrastructure (Infra)

The TAI sector provides seven high level requirements, that have to be verified or validated to prove, that AI systems comply with the AI Act. As my field is Computer Science, this is highly related. Human agency and oversight are always to keep in mind, when even just working on the idea of an AI application. It is dangerous to trick people into thinking they're talking to a real person, when they're not. Also the other requirements matter so much!
The LOR sector shares the core concepts representing the mathematical and algorithmic foundations onto which AI and its applications are built. Again highly related to Computer Science. Optimizing systems leads to better results and ideally less resources used.
The ELSA sector concerns ethical requirements, which are again highly important in the Computer Science field. Every website needs to follow certain regulations and policies (I love German bureaucracy for the imprint requirement!), so many of those requirements don't just concern AI applications.
The Infra sector states, that the infrastructure in which an AI implementation runs should facilitate its functionality in all respects. So this sector probably relates the most to the Computer Science field actually, as the requirements listed are: Data, Hardware, Architecture and Systems engineering. (TAILOR, 2024)

In a previous lecture, we explored the global "arms race" in AI development among major geopolitical powers. The European Union, however, appears to be taking a different path. Rather than competing at the same speed, it has placed strong emphasis on Responsible AI by supporting initiatives such as GPT-NL and the SRIR roadmap. Some experts argue (here and here) that this focus may cause the EU to fall behind in AI innovation.

4.2 What is your view on this? Should the EU prioritize accelerating innovation, even if it means loosening its emphasis on AI governance and ethics?

I think this should be a ITBM exclusive question, I would rather not concern myself with economic analysis. :(
I don't think the EU should loosen its emphasis on governance and ethics, instead it should advance innovation and further expand its responsible AI model. So more funding, more participation possibilities, while keeping the AI Act and trust standards. The data, that most many European firms haven't seen notable impact shouldn't be framed as innovation vs. regulation. With structures like the TAILOR map, the EU provides templates for responsible market acceleration. They target for autonomy, high-quality lawful data and safety methods. The EU should reduce the US cloud dependency, as highlighted by Accenture. Also the AI Act should be executed "innovation-friendly", so harmonized guidance, and opportunities for participation, more transparency. There should be more capacity and scale, but not less governance.

4.3 What are your key takeaways?

AI really is used everywhere!
The talking buildings were a completely new concept to me - so basically the buildings utilities like Heating/AC, Lights, Equipment and others are managed by letting the buildings communicate each with each other? Amazing!

I also really liked her example for understanding Theory of Mind:

- first "level": There is an empty vase.
- second "level": Other people recognize, that the vase is empty.
- third "level": Other people know that I know that the vase is empty.
- so on and so forth...

4.4 Will you be applying these concepts in your future profession? If so, how?

I can imagine I will get to participate in similar projects as presented in this guest lecture! They sound very interesting and there is for sure a lot to learn!

When I get back to Berlin, I will for sure tell my current supervisor about this, and maybe the Free University will copy Hanze's innovative talking buildings!

5 Week 5.2 - Trustworthy AI #2

5.1 Assignment on ALTAI. Select a specific data-driven or AI-based application in your professional field.

This may not 100% fit my professional field, but I have decided on Smart Traffic Cameras for Speeding Detection. It is important for the public sector and still very much related with Computer Science. (I contemplated between this one and AI Chatbots for Costumer Support for a long time, but someone else from Aiono will pick AI Chatbots for sure and I imagine it to be quite boring reading 20 logbooks with the same examples, so we are all profiting off my choice! Also I held a presentation about AI camera supervision (and conspiracy theories) in my first year ethics class.

5.2 Then, make an assessment of the ethical aspects concerning the application. Use the assessment list for trustworthy AI (ALTAI) as discussed in the lectures.

Starting with the first requirement **Human Agency and Oversight**, Smart Traffic Camera Systems should not operate without human supervision. Officers must verify that detected violations are legitimate before they can issue a fine. Even after a fine verified by the system and an officer, the normal appeals process is still applicable. Keeping the "normal" proceedings is important to uphold the public's trust in automated law enforcement.

Due to the second requirement, **Technical Robustness and Safety**, the AI must function accurately in all weather, lighting and traffic conditions. There should be no possibilities of missed violations or even false positives. If the camera is not able to reach the same results during day/night, or bright sun/heavy rain/snow, it is not applicable. To ensure its function, the system should be regularly calibrated and tested and fallback procedures are necessary.

Since the cameras collect personal data like license plates, **Privacy and Data Governance** are very important. The collected data must be minimal and highly protected. Footage should be stored securely, anonymized when possible, and deleted once no longer needed. In my first week at work, the Firewall Team showed us www.shodan.io, where you can easily gain access to unsecured private surveillance cameras. I think the amount of connections in Berlin has reduced since we first looked at it, so that's nice.

Transparency is an important factor in such systems. It must be available to the public, where such AI-based cameras are in use. Also an explanation about the detection and processing of violations should be provided. Authorities must disclose their responsibility for the system's operation and explanation, especially when an error occurs.

With the **Diversity, Non-discrimination and Fairness** requirement, it is important that the automated speeding detection identifies vehicles and not people. It must be tested to avoid technical bias, to make sure that no license plates will be misread or vehicle models are more targeted than others. A few years ago there was a case, where a couple got a fine for driving in the bus line, but in reality the camera identified the shirt print of a woman crossing the street as their license plate. It is hard to imagine, that there was a person actually verifying that claim (BBC, 2021).

In regards to the **Societal and Environmental Well-being**, automated speeding detection can improve road safety and reduce accidents noticeably. However, excessive and poorly communicated surveillance undermines the public trust and creates a perception of constant monitoring.

The last ALTAI requirement is **Accountability**. It must be clearly documented who is responsible for the system's operation, maintenance and error correction. Transparent reporting and complaint channels are necessary to ensure accountability throughout the systems lifecycle.

5.3 Of course, not every item on the ALTAI is of equal relevance. Show that you can identify the most relevant ethical items for your selected application.

The most relevant ethical items for my chosen application, Smart Traffic Cameras (for Speeding Detection), are in my understanding:

Accuracy and Robustness

Because it is very important to prevent false fines and forcing the fined person to then go through the appeal process to fix the systems mistakes. Also missing violations due to weather conditions will probably invite people to more reckless driving in already bad conditions.

Privacy and Data Governance

Because personal data is captured continuously and must be handled very securely. Recently, a website sparked controversy in San Francisco, because using their public data feed, the website live tracked the cities "parking cops" (Rogers, 2025). While this seems like an act of public justice, such live data feed have to be super restricted. They are prone to attacks and can never give away more than data than usually planned.

Accountability

Because there must be clear responsibility for system outcomes. There is always a big discussion about who's taking responsibility for AI systems, not just for this particular system. Smart traffic cameras definitely belong into the AI Acts "high risk" category, because they are used by law enforcement and affect citizens' rights.

High risk systems should be treated e´with even more respect than limited and low risk systems, so they therefore require very strict accuracy, oversight and compliance. All the ALTAI requirements are important when accessing this and similar applications.

5.4 What are your key takeaways?

The ALTAI requirements are:

1. human agency and oversight
2. technical robustness and safety
3. privacy and data governance
4. transparency
5. diversity, non-discrimination, and fairness
6. societal and environmental well-being
7. accountability.

Also a key takeaway of that lesson was the data privacy argument on why you should care about your data: If your healthcare suddenly doubles, because your provider got the data, that you smoke from *somewhere*, that's a problem. It's a really easily comprehensible example.

5.5 Will you be applying these concepts in your future profession? If so, how?

I think so! I got very excited when we started ALTAI two lectures prior, because when asked, what our expectations for this minor are, I think I was the only person mentioning AI ethics and Responsible AI, etc.? I think it is very nice that there are such requirements, especially because I feel like they are treated as a second priority in uni. For most project ideas/implementations/presentations the sheer usability always seems more important than the actual societal aspects. I don't think any project me or my classmates have ever implemented concerned itself in depth with those requirements.

6 Week 6.2 - Guest lecture on legal aspects of AI by Trix Mulder

6.1 The AI act recognizes a number of categories of sensitivity. Argue which category your AI application falls into.

As my AI application, I chose the product we're working on with Aiono: a voice chatbot for an ecommerce company, that takes new orders, answers questions about the order status and can respond to frequently asked questions.
I would classify the application as Limited Risk: "[...] includes AI systems with a risk of manipulation or deceit, e.g. chatbots or emotion recognition systems. Humans must be informed about their interaction with the AI." (trail, 2024)

When answering the phone, our agent, Robin, introduces himself, naming the company and pointing out, that the caller is talking to an AI. There is the risk of manipulation or deceit, because Robin could be documenting wrong order amounts, or supplying customers with wrong information, for example about opening hours. To retrace this, every call is listed, so the companies employees can easily look at the placed order and call summary, but can also read the whole transcript and if necessary contact the customer through the phone number or, if provided, even email.

6.2 Are there any issues around copyright or ownership of data in your project assignment?

I don't think that they are, because our project is just similar to their already existing agents for other companies. So for example copyright of the agents voice shouldn't be an issue, since it isn't with the existing agents.
A notable thing here is, I think, that our company was previously called Alex AI and only after our initial meeting with them did they have to change their name. They received a "note" (I'd guess a cease and desist) from Amazon, who famously own Alexa, their Smart Home Voice. While I did not think about the similarity, I am also not the person investing money in this, so I am surprised they did not receive anything earlier. They are selling their new brand identity very well though, if I hadn't known, I would've never guessed, that they made it up within a week.

6.3 What is your personal opinion on the AI act?

As I am living in student housing (lucky me), I was able to discuss this with a friend of mine, who is originally studying law in Sweden and has been going insane over her essay on the AI act over the past weeks.
She said that it is awful and everything but coherent from a law student perspective. But I am not looking from a law student perspective, so I do think, that the AI act is very important. My friend all of the articles kind of say the same thing and she really doesn't like the loopholes for unacceptable risk applications. To me it seems (maybe my law professor from last year put that thought into my head) that the language is so vague, that it just allows companies with "good" lawyers to argue their way out of consequences, which smaller companies most likely won't be able to (for example Article 15.1 "appropriate", which we talked about in the final guest lecture").
Overall I do think, that it is still important to set those regulations. It's like that crossing Grote Markt with a bike is illegal: the police will not fine it, but the sheer fact that it is illegal, makes significantly less people do it (my source for this is Ruben, maybe he was lying (also the AI Act breaches get fined)).

6.3.1 Does the law go far enough, in your opinion?

I think I am not confident in my own evaluation of a whole EU law, so I just researched widely discussed shortcomings of the AI Act: The five main flaws are:

- **Loopholes in prohibitions:** bans of unacceptable risk applications are weakened by broad exceptions, so for example, law enforcement is still allowed to used real-time face recoginition, undermining the ban.
- **Self-assessment weakens protections:** many requirements apply only to high risk apps, but companies are accessing their applications risk level by themselves. This grants them the power to avoid regulations.
- **Weak standards for fundamental rights impacts assessments (FRIAs):** the Act requires high risk systems to list potential impacts on fundamental rights, but doesn't state, if they are to proactively prevent. Apparently the requirement to consult extrenal stakeholders has been removed.
- **Broad national-security exemption:** systems used exclusively for national security purposes are excused from the AI Acts safeguards. This means governments could deploy prohibited AI without applying the AI Acts requirements *legally*.
- **Limited public participation and transparency in implementation:** there is no guaranteed meaningful participation in implementation or oversight. Complaint mechanisms exist for individuals, but NGOs can only represent cases, when consumer rights are involved.

So while the AI Act is an important development, it does not sufficiently protect the public. It gives too much weight to industry interest and law enforcement uses of AI, leaving major gaps in transparency and accountability. (ECNL, 2024)

6.3.2 Do you think the law will be effective?

I think the law will be effective, because there are a lot of very smart people working on them and I think declaring it as ineffective, even after reading the *whole* thing, would be a crazy presumption. Laws are always effective, because they provide a legal ground for assessments of certain (in this case AI related) cases.

6.4 What are your key takeaways?

The EU AI Act categorizes AI applications into different levels, from least to most harmful:

Figure 3: (Christakis & Karanthasis, 2024)/slide 18 from the slides

Examples for each levels could be:

- Low Risk: Gaming
- Limited Risk: Facial Recognition
- High Risk: Smart Cities, Border Control
- Unacceptable Risk: Social Scoring

The requirements appear to be very strict and there are a lot of them. The lecturer recommended all of us to venture into AI Law, as it is a ever evolving field.

6.5 Will you be applying these concepts in your future profession? If so, how?

The EU AI Act will definitely have a lot of impact on my future profession. As every company is just jumping on to the hype, I don't think there will ever be a workplace, where I won't have to concern myself with it. Also considering, that in the dual study program, the fifth semester is usually meant for specializing the students interest, what real-life project they wanna base their bachelor thesis on and then later work at in the company, I am probably gonna end up with an AI related topic.
The recent law classes/lectures I had in my whole Bachelors have very much piqued my interest, so I could imagine even going further into the topic. Just the combined subjectiveness of law and the constant progress in the AI field may be overwhelming.

7 Week 7.1 - Guest lecture on prompt engineering by Jos Punter

7.1 Which prompt engineering techniques did you find most useful and why?

I think the most logical technique he presented was *Instruction and Context Prompting*. While I do like to think, that especially people my age are very well able to use an AI chatbot sufficiently, I have indeed encountered too many classmates, doing exactly what was shown as the first bad example: uploading a file into ChatGPT and prompting it with "Summarize this/What's wrong with this/Give me feedback". When I talked about this to a fellow classmate the next day, saying the first technique was basically just explaining, that to get more concrete results, you have to get more concrete inputs (and that people who didn't do that before, probably still won't do it now), the answer I received was like: "Right, they tell you to do these giant prompts with so much structure and details, but you don't need all that." So basically proving my point, while totally talking past each other (weird conversation)? Prompting way too loosely and without including the aspects you want to focus on or the goal of your question/project you're asking about will just result in either the AI asking you to clarify or clarifying by yourself, because you're not satisfied with the result.

7.2 Give an example of a prompt that you have improved by applying prompt engineering techniques. Show the results before and after and explain what has improved in the results.

I try to limit my recreational ChatGPT to a minimum, so I will use an example from the From Data to Value project. To analyse the impact of bank holidays on the ACLO fitness use, I obviously need to provide bank holidays somewhere. The prompt I used was:

```
generate me a csv with the dutch bank holidays
between august 2017 and december 2021

title: Date,Name
DD-MM-YYYY,Occasion

also dont do ;

This is an example for 2015, but Brazil

Date;Name
01-01-2015;NewYearsDay
17-02-2015;Carnaval
18-02-2015;Carnaval
03-04-2015;Good Friday
21-04-2015;Tiradentes
01-05-2015;Labour Day
04-06-2015;CorpusChristi
09-07-2015;StateRebellionDay
07-09-2015;Independence Day
12-10-2015;LadyOfAparecida
02-11-2015;All souls
20-11-2015;ZumbiDosPalmares
24-12-2015;Christmas
Eve 25-12-2015;Christmas
31-12-2015;New Years Eve
```

The more intuitive prompt would've for sure been:

```
give me the dutch bank holidays
between august 2017 and demcember 2021
```

But in the prompt I used instead, I start with clarifying, that I want to receive a .csv file, not a codeblock in the chat. I provide my preferred structure, so I can immediately work with the generated file and don't have to adjust my own program to the AI deliverables. I sent the data sample from class, because that's basically what I want to generate, but I clarify, that I still want the format I set before, so commas instead of semicolons.

7.3 Read this blog on green prompting. What do you think of this concept?

I think this concept basically just combines mine and my classmates thoughts of the prior mentioned conversation. Your prompt should be as short as possible while still being as long as necessary (in German that saying rhymes, which makes it sound much wiser: so kurz wie möglich, so lang wie nötig). This concept makes so much sense, that it should even be considered common sense. But I fear it's not.
I think it is nice to appeal to the public, to rethink their usage, not only in a quantitative matter but also a qualitative one. I think such articles can leave a bitter aftertaste sometimes, because if you look at the big picture, the amount of hello's and thank you's to ChatGPT are so insignificant compared to the energy consumption of big companies. As I would like to support my decision but don't want to start a big research project right now, I did end up asking ChatGPT for confirmation by providing rough numbers of energy use.:

```
TL;DR
Yes, small messages do have an ecological impact, but it's minuscule compared to:

    - Training large AI models
    - Running them continuously for millions of users
    - Or even everyday digital activities like streaming a few minutes of video

If you want to reduce AI-related impact, using models efficiently (avoiding
redundant or repetitive prompts) matters more than skipping polite words.
```

I think it's maybe quite an ironic article to link to this lecture, since a big part of the class was all of us pasting the same prompt into our chosen chatbot, which equals to about 240 prompts in one lecture, I think?

7.4 What are your key takeaways?

My biggest takeaway from this guest lecture was, that I should maybe dive into the field of AI consulting. Also my biggest takeaway from this logbook is, that I am a very pessimistic person and don't know enough about english comma placement.

I hadn't used Suno before this lecture, so I did really like that part. As a real takeaway though, I took that artists anywhere should be scared for their profession, because why is it so accessible to produce AI music? While procrastinating uni work later that day, I created a song about my friends here, creating the lyrics with the tiniest information about everyone and then a few jokes about the dutch weather and even more jokes about ssh student housing and they loved it? I was kind of disappointed, how the lecturers version with the pro version sounded so much better than the free version, but they wouldn't have known.

7.5 Will you be applying these concepts in your future profession? If so, how?

I will be applying the concepts we've learned in my future profession, because I feel like CompSci is already a very AI dependent industry. I think it is very helpful to use it, when working with a language and program you are already familiar with, so you can properly understand what you're doing and detect the AI's errors maybe even before pasting them into your program.

I will also be applying these concepts in my current profession - being a student, because if I have to do another group work, where my partner just generates loads and loads of code, without understanding or documenting the tiniest bit... Also if I have to witness another person pasting my project reports into ChatGPT "for clearer structure and nicer wording", I will crash out, kindly.

8 Week 7.2 - Guest lecture on AI Standardization by Willy Tadema & Theresa Marschall

In the guest lecture, we saw a projected timeline of the AI Act coming into force.

8.1 Do you think this timeline is realistic, based on what you have seen so far in practice? For example, in the company you're working with for your group project, or a different company in your field.

I also discussed this question with my aspiring lawyer friend and as lawyers usually say: it depends. Overall the AI Act timeline seems very ambitious, since it sets very strict compliance and documentation requirements. This will pose a struggle especially to smaller companies, because adapting will take not only time but also understanding of the requirements. My first thought was, that this could probably be a problem for Aiono (my project company), as they are only six employees, if i remember correctly. But then I recalled, that they are working very closely with Problicity, who is to my understanding a significantly bigger company, that should have enough money and employees to concern itself with the requirements.
Larger companies are definitely better positioned to adapt to the AI Act, as they should have existing compliance frameworks.

8.2 What are your key takeaways?

I really liked their explanation of what the RDI works on and the standardization examples, because as a European, who has never traveled significantly far, standardized power outlets for example are barely on my mind. I really appreciate, that they talked about the Netherlands past biased algorithms, and showed us the example with the categories, explaining why they cannot be used in fraud detection.
I also highlighted Article 15.1 (Future of Life Institute, 2025) in my notes, which cites:

> `High-risk AI systems shall be designed and developed in such a way that they achieve an appropriate level of accuracy, robustness, and cybersecurity, and that they perform consistently in those respects throughout their lifecycle.`

While I understand the "confusion" about the word *appropriate*, as far as I'm aware, most EU regulations are phrased like that. And since there is a reason, high risk systems are considered high risk (and we did talk about examples not just in this lecture but also in the AI Law lecture), there should definitely be a standard established there.

I think my classmates takes "If my racist fraud detection works and actually detects more fraud, shouldn't it be allowed" and "The EU regulates AI too much, we're falling behind the US" we're especially weird this lecture. Ruben said the first one is "such a german take", but I'd rather classify it as a white male take, so he can keep Titus to himself, we don't claim him.

8.3 Will you be applying these concepts in your future profession? If so, how?

I already talked about the AI Act for this question in the AI law chapter. I think that standardization, not just in an AI context will definitely matter in every CompSci related profession, because the whole point of technology is, that it works in different contexts/countries/etc., right? So for example many projects from the Free Universities Identity Management are also used by different german universities, since they do have the same use cases and it is one of Germany's biggest/most funded public universities. To make that possible, a certain standard has to be met. It probably won't take too long for them to also work on AI related applications here.
The two lectures did say though, that working on AI standards equals reading a lot, discussing a lot and writing a lot, which doesn't sound too appealing.

References

Andreessen, M. (2023). Why AI Will Save the World [https://social.ayjay.org/uploads/2023/bda5231720.pdf]. *a16z*.

BBC. (2021). Bus lane camera mistakes woman's sweater for number plate [https://www.bbc.com/news/uk-england-somerset-58959930]. *BBC*.

Bell, J. (2025). Stop Calling it Hallucination LLMs are Working Exactly as Designed. [https://cacm.acm.org/blogcacm/smarter-prompts-for-a-more-sustainable-future/]. *Medium*.

Bhaskar Chakravorti, A. B., & Chaturvedi, R. S. (2019). Which Countries Are Leading the Data Economy? *Harvard Business Review*.

Bhattacharyya, S. (2023). Global Digital divide, Censored Digital Freedom and Bias in Generative AI models [https://medium.com/@bhattacharyya.shilpi.sbu/global-digital-divide-censored-digital-freedom-and-bias-in-generative-ai-models-22a04b94f5a6]. *Medium*.

Christakis, T., & Karanthasis, T. (2024). AI EU Act: A Risk-Based Approach [https://ai-regulation.com/visualisation-pyramid/]. *AI-Regulation.com*.

Consultancy.eu. (2024). Europe lagging behind US in AI arms race, warns Accenture report [https://www.consultancy.eu/news/12083/europe-behind-us-in-ai-arms-race-accenture-report]. *Consultancy.eu*.

Csernatoni, R. (2025). The EU's AI Power Play: Between Deregulation and Innovation [https://carnegieendowment.org/research/2025/05/the-eus-ai-power-play-between-deregulation-and-innovation?lang=en]. *Carnegie Europe*.

ECNL. (2024). Packed with loopholes: Why the AI Act fails to protect civic space and the rule of law [https://ecnl.org/news/packed-loopholes-why-ai-act-fails-protect-civic-space-and-rule-law]. *European Center for Not-for-Profit Law*.

European Comission. (2020). Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment [https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment]. *European Comission*.

Future of Life Institute. (2025). Article 15: Accuracy, robustness and cybersecurity [https://artificialintelligenceact.eu/article/15/]. *EU Artificial Intelligence Act*.

Future of Privacy Forum. (2025). EU AI Act [https://fpf.org/wp-content/uploads/2025/04/FPF_EU_AI_Act_Timeline_R4_-_2025_Update.pdf]. *Future of Privacy Forum*.

Grider, R., & Wilensky, U. (2015). NetLogo Paths model [http://ccl.northwestern.edu/netlogo/models/Paths]. *Center for Connected Learning and Computer-Based Modeling*.

IBM. (2025). What is AI ethics? [https://www.ibm.com/think/topics/ai-ethics]. *IBM*.

Kumar, A. (2025). Types of AI Explained [https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/types-of-artificial-intelligence]. *simplilearn*.

Molhoek, M. (2025). Federated learning: get to know privacy-preserving data analysis [https://www.tno.nl/en/technology-science/technologies/federated-learning/]. *TNO*.

Mouriquand, D. (2025). What is the Pentagon Pizza theory eating away at the internet? [https://www.euronews.com/culture/2025/06/06/what-is-the-pentagon-pizza-theory-eating-away-at-the-internet]. *euronews.*

Rogers, R. (2025). For One Glorious Morning, a Website Saved San Francisco From Parking Tickets [https://www.wired.com/story/san-francisco-find-my-parking-cops/]. *WIRED*.

TAILOR. (2024). Strategic Research & Innovation Roadmap of Trustworthy AI [https://tailor-network.eu/wp-content/uploads/2025/05/TAILOR-Roadmap-Version-2-1.pdf]. *TAILOR*.

trail. (2024). The four risk classes of the EU AI Act [https://www.trail-ml.com/blog/eu-ai-act-how-risk-is-classified]. *Official Journal of the EU*.

UN Global Compact. (2025). Artificial Intelligence and the Sustainable Development Goals: Operationalizing technology for a sustainable future [https://unglobalcompact.org/compactjournal/artificial-intelligence-and-sustainable-development-goals-operationalizing]. *UN Global Impact*.

Wilensky, U. (1997). NetLogo Party model [http://ccl.northwestern.edu/netlogo/models/Party]. *Center for Connected Learning and Computer-Based Modeling*.

Williams, A. (2025). Smarter Prompts for a More Sustainable Future? [https://cacm.acm.org/blogcacm/smarter-prompts-for-a-more-sustainable-future/]. *Communications ACM*.