

## Protokoll 29

---

*Guido Zurstiege & Victoria Smith Ekstrand*

### **It's terrible what you are doing – how much does it cost?** (Guido Zurstiege)

Europas größte Forschungsk Kooperation auf dem Gebiet der Künstlichen Intelligenz, das Cyber Valley, spielt sich hier in Tübingen überwiegend oben auf dem Berg ab – dort, wo in den vergangenen Jahren viele neue Bürogebäude aus dem Boden geschossen sind. Unten im Tal befinden sich weiterhin die Geistes- und Sozialwissenschaften sowie neben Teilen der Verwaltung auch das Rektorat der Universität. Von dort kam bereits kurze Zeit nachdem ChatGPT im November des vergangenen Jahres für die Öffentlichkeit frei zugänglich geworden war per Rundschreiben eine an alle Angehörigen der Universität adressierte Nachricht der Rektorin – darin: die aktuelle rechtliche Beurteilung jener KI, die in den Wochen zuvor wie ein apokalyptischer Reiter der Singularität die Schlagzeilen gestürmt hatte. Von ChatGPT erstellte Texte dürften nicht verwendet werden, auch nicht bei Offenlegung der Quelle. So ist es also nicht ganz einfach, hier auf dem Talboden des Cyber Valley darüber zu reflektieren, woran oben an dessen sonnenbeschienenem Kranz so intensiv geforscht wird. Mein Bericht muss daher wohl oder übel im Modus der "kalten Faszination" (Siegfried, J. Schmidt) erfolgen.

Ungeachtet der klaren Ansage spielen freilich alle mit der Quatschmaschine, füttern sie unaufhörlich mit ihren Fragen, so dass sie weiter wächst und gedeiht. Für einen Vortrag hat mein Freund T. mit Hilfe von ChatGPT seine wissenschaftliche Vita zusammenfassen lassen. Es musste schnell gehen. Vor allem hatte ihn aber die Neugier getrieben – wie uns alle. Mit dem Ergebnis war er recht zufrieden. "Vanity Chatting – gibt es den Begriff schon?" denke ich und versuche es dann auch einmal, obwohl das natürlich ein bisschen peinlich ist. In meinem von der KI generierten alter ego erkenne ich mich kaum wieder. Bald wird es "search engine optimization" auch für chat engines (CEO) geben, sagt H., dem ich von meinem missglückten Versuch berichte. Ich sollte unbedingt an meiner Netz-Präsenz arbeiten.

Beim zweiten Anlauf wähle ich eine berühmte Persönlichkeit des Fachs: "Wer ist Elisabeth Noelle-Neumann und was war ihre Rolle zur Zeit des Nationalsozialismus?" will ich wissen und erfahre, dass die Grande Dame der Demoskopie in Deutschland Mitglied der Widerstandsgruppe "Weiße Rose" war. Moment mal? War sie nicht mit Albert Speer irgendwie befreundet gewesen? Hatte sie in ihrer Dissertation (1940, S. 63) nicht behauptet, die Juden hätten seit 1933 einen Großteil des geistigen Lebens in den USA monopolisiert, um "ihre demagogischen Fähigkeiten auf die Deutschlandhetze" zu richten? Auch ein Fall für CE-O.

Man muss schon wissen, wonach man fragt, sonst sieht man schnell alt aus. Oder man fragt nach etwas, das man nicht wissen kann und daher auch nicht unbedingt wissen muss. Das tue ich und ziele auf eine feine Masche im Geflecht des Geistes, die Intuition. Ohne sie träte keine Musikerin den richtigen Ton und kein Fußballer das Tor, keine Chirurgin setzte den richtigen Schnitt, kein Maler den passenden Strich. Und die KI? "Frühling" schreibe ich, sonst nichts. Dann bekomme ich eine durchnummerierte Liste mit "Aspekten" der Jahreszeit, die mich hoffentlich inspiriert, heißt es in der Antwort: Blütenpracht, wärmeres Wetter, längere Tage, frische Luft, neue Anfänge – solche Dinge. Ich bitte ChatGPT danach um ein paar Zeilen Programmiercode in R, für die graphische Darstellung des durchschnittlichen Temperaturverlaufs in Deutschland. Die KI hat auch darauf eine gute Antwort.

Der Computer „gleicht einem Menschen weniger als eine Amöbe; dennoch ähnelt er dem Gehirn wie keine andere Maschine zuvor. Er ist nah genug, um die Menschen erzittern zu lassen". So hat der Mathematiker James R. Newman (1956, p. 2067) bereits Mitte der fünfziger Jahre das Kopf-an-Kopf-Rennen zwischen Gehirn und Computer beschrieben. Wo die Pioniere der KI-Forschung das Gehirn, ins Zentrum rückten, da geht das maschinelle Lernen aber tatsächlich von einer Tabula rasa aus. Nicht der Geist steht am Anfang, sondern ein ganz und gar leeres Blatt, das Stück für Stück mit Mustern beschrieben wird. So haben wir die Maschinen auf einen Beobachtungspunkt zweiter Ordnung gesetzt, den wir mangels Talent selbst niemals werden einnehmen können.

Eine wirklich intelligente Maschine würde freilich nachfragen "Frühling? Was meinst Du?" ChatGPT tut das nicht. Überhaupt ist eines der auffälligsten Merkmale der KI die Gewissheit, mit der sie antwortet, ohne je die Contenance zu verlieren. Der Grund dafür liegt im verdrängten Unbewussten des digitalen Zeitalters, der Materialität. Es wiederholt sich ein bekanntes Muster: Wo sich die Schulkinder hierzulande auf „das Ende der Kreidezeit“ vorbereiten, befinden sich die des globalen Südens weiterhin in einer humanitären Steinzeit, die sie dazu verdammt, jene Ressourcen, die wir für die Produktion unserer glän-

zenden Gadgets benötigen, mit bloßen Händen aus der Erde zu kratzen. Genauso ist es auf den vorgelagerten Produktionsstufen der KI. Sie erspart uns den Blick in den Abgrund der menschlichen Intelligenz nur deshalb, weil sich andere jeden Tag für uns in das dunkelste Schwarz der Seelengrube hinab begeben, wo sie gegen einen Hungerlohn unvorstellbare Bilder und Texte zu sehen bekommen.

Das Netz hat Wissen und Unterhaltung in ungeahnter Fülle verfügbar gemacht und sich zugleich als Triebkraft eines neuen, rückwärtsgewandten Tribalismus erwiesen. Noch nie standen den Menschen so viele Möglichkeiten der Information zur Verfügung wie heute. Paradoxerweise waren aber auch die Chancen für alle Schurken dieser Welt niemals besser, ihre mit Propaganda und Desinformation verdienten Dividenden einzustreichen. Das digitale Zeitalter beherbergt eine prekäre epistemische Gemeinschaft. Ob Bild, Text oder gesprochene Sprache, schon lange ist das alles offen für die Manipulation der Zeichen. So trägt auch die KI weiter zu einer geradezu epochalen Generalisierung des Zweifels bei. Trump im Ringkampf mit New Yorker Ordnungshütern, der Papst in weißer Edeldaune, beides war zwar schnell als sweet Fake entlarvt, auf den Titelseiten der Qualitätspresse entfaltet es aber dennoch eine die Glaubwürdigkeit der Kommunikation zersetzende Wirkung, weil es so täuschend echt aussieht.

Nun wird der Ruf nach einem Moratorium laut – ein halbes Jahr Entwicklungspause, um die Risiken für die Menschheit einzuschätzen. Nach jahrzehntelanger Forschung und milliardenschweren Investitionsprogrammen wirkt alleine dies wie ein Akt der Verzweiflung. Oder ist dies wie die amüsanten Bilder Trumps und des Papstes Teil einer raffinierten Vermarktungsstrategie, die man sich von den persuasiven Industrien des vergangenen Jahrhunderts abgeschaut hat? Eine der meistgelesenen Polemiken gegen die "captains of consciousness" (St. Ewen) jener Zeit war Vance Packards „Die geheimen Verführer“. Es attackierte die damaligen Expert:innen der Manipulation, die in ihren psychotechnischen Laboren daran arbeiteten, den freien Willen des Menschen durch subliminale Kommunikation zu unterjochen. Einer der bekanntesten dieser von Packard kurz „Tiefenheinis“ genannten Strategen im Reich der Wünsche war der aus Österreich stammende Motivforscher Ernst Dichter. Nach der Publikation von Packards Buch, so berichtete dieser später einmal, habe er Briefe über Briefe von den schockierten Leser:innen erhalten. Alle mit dem gleichen Inhalt: „Es ist furchtbar, was Sie tun, Dr. Dichter! Wieviel kostet es?“ (Dichter cf. Norris, 1984, p. 34) Die Kritik an den mächtigen Medien, die ihr Publikum in den Zustand ohnmächtiger Empfänger persuasiver Botschaften versetzten, hat seit jeher nicht nur den Widerstand gegen sie befeuert, sondern auch die Hoffnungen all jener, die massive Steuerungsinteressen verfolgen und danach trachten, sie mit Hilfe der jeweils neuen Medien durchzusetzen. Als sich Google am Anfang seiner steilen Laufbahn verpflichtete,

nichts Böses zu tun, da wuchsen also die Angst und die Hoffnung miteinander – nur in verschiedenen Herzen.

Verfolgt man die aktuelle Debatten über die Macht der KI, sieht man, dass die Ingenieure des Digitalen Zeitalters eine noch größere Aura der Allmacht umgibt als deren Print- und Rundfunkvorgänger. In gewisser Hinsicht sind die aktuellen Schlagzeilen eine Form jener selbsterfüllenden Prophezeiung, die der französische Philosoph Jaques Ellul in seiner viel zitierten Studie über die Entstehung und Lenkung der öffentlichen Meinung als „Vorpropaganda“ (2021 [1962]) bezeichnet hat: eine Form der strategischen Kommunikation, die sich qua Behauptung ihre eigenen Wirkungsvoraussetzung schafft.

#### Literatur:

Newman, J. R. (1956). *The world of mathematics*. London: George Allen and Unwin.

Noelle, E. (1940). Amerikanische Massenbefragungen über Politik und Presse. Frankfurt/M.: Verlag Moritz Diesterweg.

Norris, E. (1984). Ernest Dichter. *Success. The Magazine for Achievers*(511), 33-42.

Ellul, J. (2021 [1962]). *Propaganda: Wie die öffentliche Meinung entsteht und geformt wird*. Frankfurt a.M.: Westend-Verlag.

#### “GPT Justice” (Victoria Smith Ekstrand)

*“Where architectures of code change the constraints of law, they in effect displace values in the law. Lawmakers will then have to decide whether to reinforce these existing values, or to allow the change to occur. ... [Code] can be made to constitute values that resonate with our tradition, just as it can be made to reflect values inconsistent with our tradition.”*

Lawrence Lessig, “The Law of the Horse: What Cyberlaw Might Teach,” 1999.

More than 20 years ago -- long before smartphones, social media, and AI -- Harvard Law professor Larry Lessig told us that “code is law.”

As a then 38-year-old law professor at Stanford University, Lessig published his famous axiom in several articles and in his best-selling book, *Code and Other Laws of Cyberspace*. He argued not only that a legal disturbance was upon us, but a consequential shift had already occurred. He challenged a federal judge who had declared that there was “no more a law of cyberspace than there was a ‘Law of the Horse,’”

and warned the neoliberal tech evangelists of the time that threats would come from within. These were the men (and they were largely men) who subscribed to cyberlibertarian John Perry Barlow's creed: "Governments of the Industrial World, you weary giants of flesh and steel, I come from Cyberspace, the new home of Mind. On behalf of the future, I ask you of the past to leave us alone. You are not welcome among us. You have no sovereignty where we gather."

To them and to us, Lessig issued his warnings: In the digital era, threats or constraints would come from three sources in addition to government: social norms, the market, and architecture – and more precisely, the architecture of code.

Our task, Lessig told us, was to identify the values we wished to preserve and protect in cyberspace, both substantive and structural. The structure of code embedded certain values as well as a type of regulation, and we needed to pay heed to this opaque double helix. While his project gained much attention among scholars, it did little to alter either the substance of the law or the direction of platform architectures. Surveillance and disinformation flourish. Platform transparency barely exists. Laws largely favor markets.

When I once complimented him on his prescience, Lessig simply lamented his role. "I don't relish being a Cassandra," he told me.

Now that Barlow's future may have arrived, I posed the question to ChatGPT: Was Lessig indeed a Cassandra? Is ChatGPT a realization of Larry Lessig's claim that code is law? It replied:

*ChatGPT is not a direct realization of Larry Lessig's claim that "code is law," as it refers specifically to the way that computer software code can act as a form of regulation in cyberspace. ChatGPT is a language model that is designed to understand and generate natural language, and it does not operate in the same way that software code does.*

*However, ChatGPT can be seen as an example of how technology can shape the way that people interact with each other and with information. ... Both ChatGPT and software code have the potential to influence how people think, communicate, and make decisions, and both can be seen as forms of "law" in their respective domains .... While ChatGPT does not operate in the same way as software code, it can be seen as a manifestation of the power of technology to shape human behavior and decision-making.*

If that is the case, what values do the programmers of ChatGPT instill? What does its code tell us about the "law" of AI? ChatGPT, not surprisingly, sees its "natural language" capabilities (as if anything about language is "natural?") as value-neutral:

*While there are no inherent values that are explicitly programmed into ChatGPT, the quality and accuracy of the information that it provides can*

*be influenced by the data that it has been trained on. ...The responses generated by ChatGPT are not intended to reflect any specific values or ideologies, but rather to provide useful and accurate information to the best of its ability based on the input text that it receives.*

I decided not to quibble. I became direct. I then simply asked if it posed a danger to society. A “both sides” pattern emerged as it had in the previous two queries: The first paragraph was “no.” The second paragraph posed possible exceptions to the first. The third cautioned against an absolute judgment either way. The conclusion could be characterized as absolution. I felt I had entered a courtroom, or at the very least, the outline of a brief to a court:

*First Paragraph: ChatGPT is a language model developed by OpenAI that is designed to facilitate communication and provide information to users. Like any technology, it has the potential to be used in both positive and negative ways, but in and of itself, it is not inherently dangerous.*

*Second paragraph: However, it is important to recognize that AI technologies like ChatGPT can have limitations and biases based on the data that they have been trained on. ...*

*Third paragraph: It's also important to note that AI technologies like ChatGPT are not intended to replace human interaction or decision-making. ...*

*Conclusion: In summary, ChatGPT itself is not dangerous, but it's important to recognize the limitations and potential biases of AI technologies and to use them in a responsible and ethical manner.*

Whatever ChatGPT’s “thoughts” on the matter, the regulatory apparatus is not waiting. In the U.S., the U.S. Federal Trade Commission, the Civil Rights Division of the U.S. Department of Justice, the Consumer Financial Protection Bureau, and the U.S. Equal Employment Opportunity Commission issued a joint statement in April 2023 outlining a commitment to enforce their respective laws and regulations to promote responsible innovation in automated systems. All four agencies had previously expressed concerns about AI.

“We already see how AI tools can turbocharge fraud and automate discrimination, and we won’t hesitate to use the full scope of our legal authorities to protect Americans from these threats,” said FTC Chair Lina Khan. “Technological advances can deliver critical innovation—but claims of innovation must not be cover for lawbreaking. There is no AI exemption to the laws on the books, and the FTC will vigorously enforce the law to combat unfair or deceptive practices or unfair methods of competition.”

In Europe, the European Parliament’s main political groups reached an agreement on its AI Act, with lawmakers addressing prohibited practices, emotion recognition, and

predictive policing. The UNESCO Recommendation on the Ethics of AI and the White House Blueprint for an AI Bill of Rights are additional guiding documents. Italy did not quibble. It made the decision to block ChatGPT altogether.

Will law chase AI, or will AI consume the law? More to the point, will justice be automated? As Tim Wu points out, the idea that systems of public ordering might compete is not new, but “what is new is the arrival of software and artificial intelligence as a major modality of public decision-making.” Speech, once primarily governed by law, is already “now governed by a hybrid-human machine system of social ordering,” Wu writes, and much of it is now driven by free speech values closer to European democracies than to U.S. free speech principles, at least on social media.

Whether the power of artificial intelligence in ChatGPT accelerates the move to a less hybrid model of “code as law” remains to be seen. What AI programs like ChatGPT offer is something closer to “enforcement capacity.” It is fast, it can scale, and it operates efficiently, as Wu also points out. More convincingly, its reasoned, nuanced “voice” seems to offer the potential for the kind of procedural fairness present in real courtrooms and still lacking or less transparent in more basic modes of social media content moderation.

The “voice” of ChatGPT -- first offering an argument against, then for, then proceeding with caution -- conjures the voices of litigators and judges asked to understand and rule on the nuances of disputes. It’s an output that feels far afield from the definitive takedown algorithms of social media.

Code as law, yes. Code as ChatGPT is potentially the “natural language” of the courtroom.

