

## Unite! Summer School, June 17-21, 2024 at TU Darmstadt / Germany

### What Computers and AI can't do.

#### Reflecting on Errors, Malfunctioning, and Limitations of Technology in the Digital Age

AI critic Hubert Dreyfuss's 1972 book "What Computers can't do" is one of the classics of the early AI debate. Dreyfuss criticized the AI of the time (the so-called symbolic AI) and insisted that it could never replicate human thought processes. He diagnosed a fundamental limitation of AI. Another prominent critic of the time, Joseph Weizenbaum, argued ethically and called for a rather careful consideration of what computers should and should not do.

Today, in the context of statistically based approaches to machine learning, limits, flaws, and failures in digital processing are discussed. The bias of AI is criticized. Errors in autonomous vehicles frighten people, and the hallucinations of Chat GPT have been intensively commented on.

While the public debate focuses mainly on what AI can do, how it could make the world a "better place," and what dangers it might pose to people and societies, the Summer School will focus on the limits, errors, and malfunctions in the digital age, thus on the power and politics of failures. The functioning of society depends on the functioning of digital technology and probably soon on AI. Thus, their failures, malfunctions, and limitations are highly relevant for contemporary societies.

We will discuss the following, interwoven themes:

#### *The Public Discourse on Errors, Limits, and Malfunctions in the Digital Age*

In the 1970s in Germany, a heated debate began about the possible effects of software errors. As early as 1968, computer scientists had discussed errors and failed software projects at an international conference. Soon, it also became a topic of public debate. In the 1990s, the vulnerability of the information society was a topic of much concern. Today, we are discussing errors, malfunctions, and the limits of AI.

There is often a public perception that technology must be flawless; on the other hand, failures and malfunctions are part of everyday life in digital societies. Errors, such as a recent system failure in British air traffic control, grab the attention for a moment but quickly fade from the discussion.

We want to discuss how the discourses on errors, malfunctions, and the limits of computers and AI have changed from the second half of the 20th century to the present. What are the public's expectations? When is an error perceived as a problem?

What role do technical errors and malfunctions play in public perception and how do they influence public opinions toward computers and AI? How do computer scientists and AI researchers act and communicate in public?

Finally, how should we publicly discuss bugs and glitches in the digital age? How do we balance fear-mongering with informed skepticism and responsible behavior? What emotions are triggered by errors, malfunctions, and limits of digital technology and AI? Fear, anger, distrust?

### *Concepts of Error*

What is considered an error or a fault? Who defines if something is an error? How has the concept of error changed historically? How is it judged whether it's a relevant or even a serious error? In the context of AI bias, it has been stressed that bias is "more than a glitch" (Broussard 2023) i.e. not an error but a feature inherent to AI. This leads to the question whether technical failures are also always social failures. Can we speak of purely technical failures at all? How are software and AI errors conceptualized in public?

### *Tackling Errors/Finding Errors/Scientific Culture of Error*

Errors have long been a topic in the history of science. The role of errors in the production of knowledge has been studied, and the productive role of errors has been emphasized.

However, the question of how computer scientists and AI researchers deal with errors is not sufficiently reflected in Science and Technology Studies. What kind of error culture has developed since the emergence of computer science? Which error culture prevails in AI research? These issues should be the subject of interdisciplinary discussion.

This also leads to highly relevant epistemological questions of the present: How can computer and AI models be verified and validated? How do users know whether decisions made by AI are correct e.g., in medical diagnostic systems? Can they even know? That is one of the most topical questions in the context of AI and the attempts to develop explainable, transparent AI.

### *Ethical and Scientific Limits of Computers and AI: What They Can't and Shouldn't Do*

The debate about the limits of AI from the 1970s and 1980s needs to be revisited today under new auspices. A historical perspective can help in assessing current developments and concerns. That includes a scientific dimension, e.g., the question of the limits of the current path of machine learning. Is the new approach still in its infancy or will it soon reach its fundamental limits? And what do current limits mean?

Moreover, the ethical debate is also being revived and needs to be addressed: Which AI applications are socially appropriate or dangerous? How should we deal with the limits and failures of AI? Is regulation the way to go? Which positions are taken by whom?

The Summer School will leave room for additional topics and other focus areas.

## **About the Summer School**

The Summer School is a cooperation of UNITE! with scholars from European universities. UNITE! is a network of universities in nine countries that connects engineering, science, and technology with the grand challenges of society. The Summer School is jointly organized by scholars from TU Darmstadt, Aarhus University, TU Graz, Nova University Lisbon, and KTH Stockholm.

The Summer School is the first event of a planned series of Summer Schools within the next years. It aims to establish an interdisciplinary dialogue between humanities, social sciences, computer science, and engineering science on the so-called Great Challenges, focusing on changing topics. The upcoming Summer School on *Errors, Malfunctioning, and Limitations of Technology in the Digital Age* will offer lectures, intense discussions in working groups, discussions of participants' research projects, and a final round table with experts, which will be open to the wider public. Participants will receive a reader with texts and materials for the seminars. The working language is English. Excursions are also planned.

Further co-organizers of this Summer School are the Interdisciplinary Studies Programme (isp) and the Research Field Information and Intelligence at TU Darmstadt.

### **Who can participate in the Summer School?**

We invite students and Ph.D. students of history, STS, philosophy, sociology, and neighboring humanities disciplines, as well as students and Ph.D. students of engineering science and computer science to participate.

### **How to Participate?**

Participants can present and discuss their research project with e.g., a lecture, a poster, an intervention, or a science slam. Should you like to participate without giving a presentation of a current research project, reading texts and active participation are expected.

The standard number of ECTS points for the Summer School is 5 - if required, other amounts can be arranged. Attendees will receive a certificate for their participation in the Summer School.

### **How to Apply?**

All applications should be submitted electronically in PDF format and should include the following:

1. Letter of Intent indicating interest in the Summer School's annual topic (max. 300 words);
2. Curriculum Vitae (max. 2 pages);
3. Abstract of a possible presentation at the Summer School (max. 450 words)

To apply, please send the documents by March, 31, 2024 as **a single** PDF file to [hessler@pg.tu-darmstadt.de](mailto:hessler@pg.tu-darmstadt.de)

### **Time and Place**

The Summer School will take place in Darmstadt on June 17-21, 2024.

**Upon request, travel funds are available.**

If you have any questions about the program or the application, please contact Prof. Dr. Martina Heßler: [hessler@pg.tu-darmstadt.de](mailto:hessler@pg.tu-darmstadt.de)