UNIVERSITÄT MANNHEIM

AI and Data Science in Fiction and Society

Seminar Kick Off

Heiko Paulheim

Goals

- Why should we read fictional literature at university at all?
  - given that we are not humanities students

- Fiction is a mirror of society
- Technology is developed in research labs
  - ...but its impact on society is seldom discussed therein

- In this seminar, we will discuss those impacts

Organization

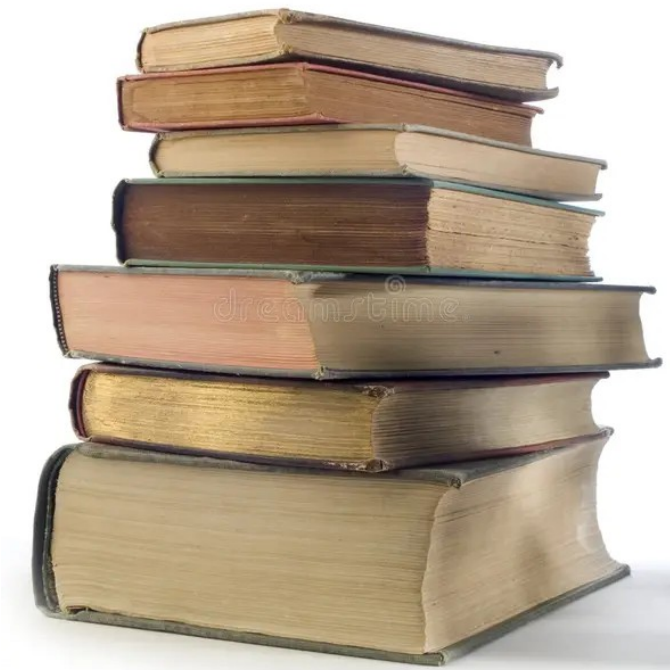
- Requirements
  - Read a novel and present it in the seminar
  - Write a seminar paper
  - Review others' seminar papers
    - it is a good idea to also read the books for the seminar papers you review

- First step
  - Pick a novel
  - If not done yet, send a ranked list to Ms. Ezgi Yilmaz
    - Until the end of Sunday, Sept. 21st
  - You are also invited to propose books **not** on the list on the Web page

Organization

- Getting your book
  - we have copies in the library
    - of books listed on the Web page, not own suggestions!
  - you may need them for some time

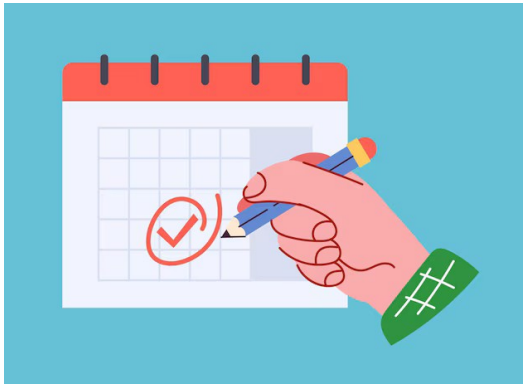
- After assignment is finished, you can borrow “your” copy
  - address the counter in the library in the castle west wing
  - please wait for the confirmation mail until you try to borrow “your” copy



Organization

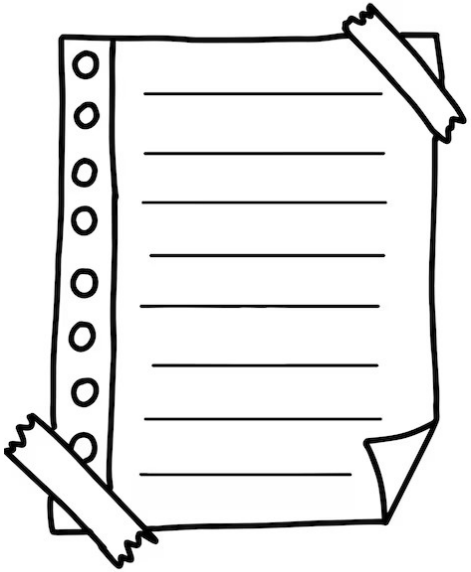
- We will use a process called “peer review”
  - widely used (and discussed) in science
  - you will review your fellow students’ seminar papers

- Timeline
  - Prepare a draft until October 19th
  - You will get two seminar papers to review
  - Submit your reviews until October 26th
- Seminar (i.e. ,presentations, discussions)
  - October 30th, November 13th, 20th, 27th
  - Participate *actively*
- Final seminar paper submission: January 18th, 2026



Preparing Your Seminar Paper

- Synopsis
  - Brief(!) summary of the novel's contents
  - Main themes, characters, plot points
- Characters
  - Main characters? What do they represent?
- Technology
  - Which technology is described?
  - What does already exist, what is still to be invented?
- Interaction of technology and society
  - Which of the characters appreciate the technology?
  - Who are rather sceptics?
- Anything else you want to discuss
- Use further literature!



Hints for Further Literature

- Book reviews
  - e.g., in newspapers
  - which aspects do they appraise? which do they criticize?
- Information on the author
  - e.g., interviews
  - did s/he comment on the book?
- Information on technology
  - e.g., research articles
- Information on ethical aspects
  - e.g., research articles
  - e.g., high level documents (EU HLEG, to come later)
- Information on legal aspects (e.g., EU AI Act)

AI Tool Usage

- You're supposed to use the DWS seminar template
  - https://www.uni-mannheim.de/dws/teaching/thesis-guidelines/

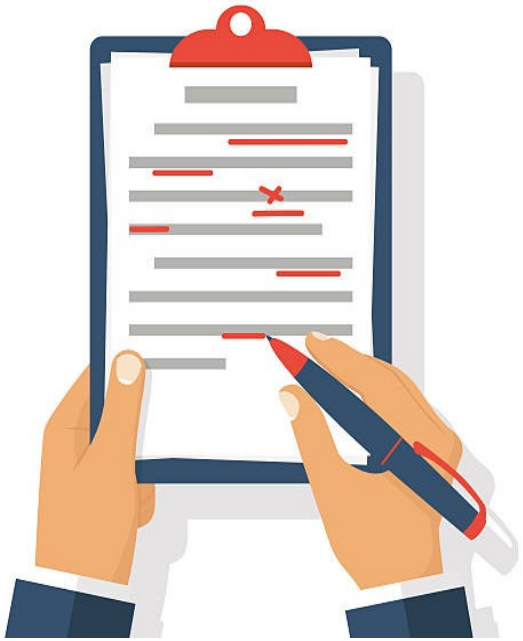
- You may use AI, but have to declare it (and rate its utility):

Declaration of Used AI Tools

Tool	Purpose	Where?	Useful?
ChatGPT	Rephrasing	Throughout	+
DeepL	Translation	Throughout	+
ResearchGPT	Summarization of related work	Sec. 2	-
Dall-E	Image generation	Figs. 2, 3	++
GPT-4	Code generation	functions.py	+
ChatGPT	Related work hallucination	Most of bibliography	++

Preparing a Peer Review

- 1st rule: be constructive!
- What you should point at
  - can you follow easily?
    is there information missing at any point?
  - are all claims well supported?
  - do you have any questions not answered?
  - aspects underrepresented
- What you should not do
  - provide general criticism (“don’t like the paper”)
  - correct every spelling mistake
  - rewrite the seminar paper



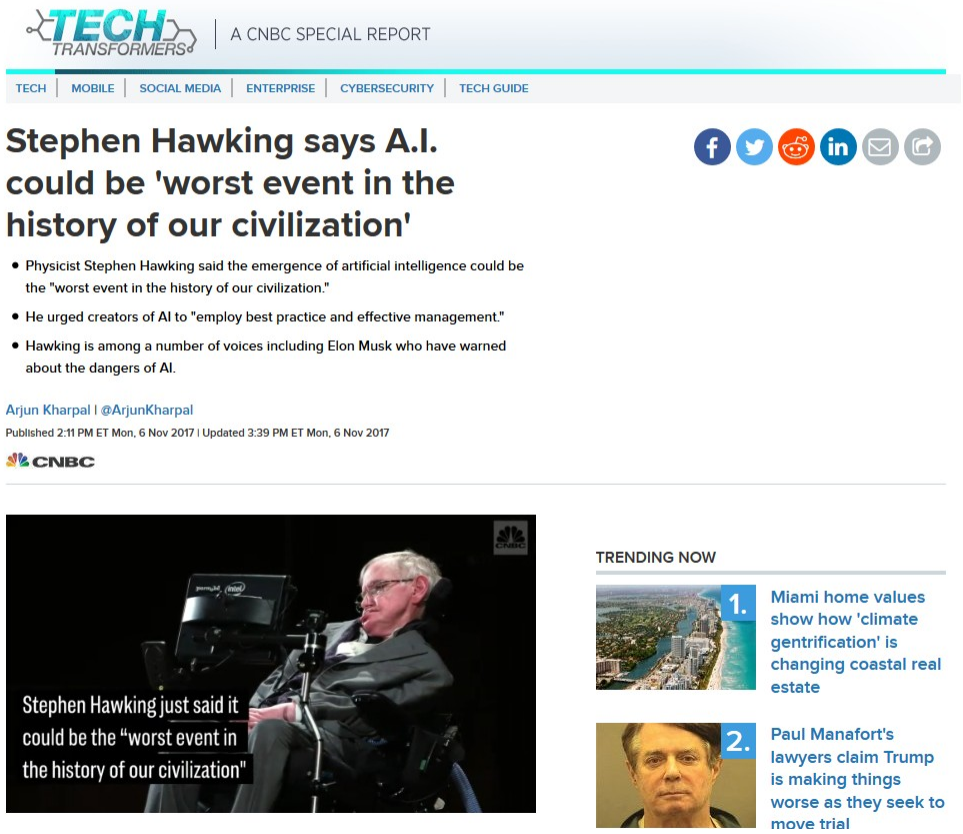
Preparing the Presentation

- Remember
  - we are here to discuss
- Therefore
  - don’t just report facts
  - raise questions
  - be provocative
- Structuring your talk
  - spend max. 10 minutes on plot and characters
    - it is OK to omit details
  - the rest should focus on discussion of technology and society

Questions?

Straw Poll (will be repeated at the end)

- Do you consider AI rather beneficial or harmful?



TECH TRANSFORMERS | A CNBC SPECIAL REPORT

TECH | MOBILE | SOCIAL MEDIA | ENTERPRISE | CYBERSECURITY | TECH GUIDE

Stephen Hawking says A.I. could be 'worst event in the history of our civilization'

- Physicist Stephen Hawking said the emergence of artificial intelligence could be the "worst event in the history of our civilization."
- He urged creators of AI to "employ best practice and effective management."
- Hawking is among a number of voices including Elon Musk who have warned about the dangers of AI.

Arjun Kharpal | @ArjunKharpal

Published 2:11 PM ET Mon, 6 Nov 2017 | Updated 3:39 PM ET Mon, 6 Nov 2017

CNBC

TRENDING NOW

1. Miami home values show how 'climate gentrification' is changing coastal real estate

2. Paul Manafort's lawyers claim Trump is making things worse as they seek to move trial

Benevolent vs. Malevolent AI

AI 2027 [1]

Summary Research About

Daniel Kokotajlo, Scott Alexander, Thomas Larsen, Eli Lifland, Romeo Dean

We predict that the impact of superhuman AI over the next decade will be enormous, exceeding that of the Industrial Revolution.

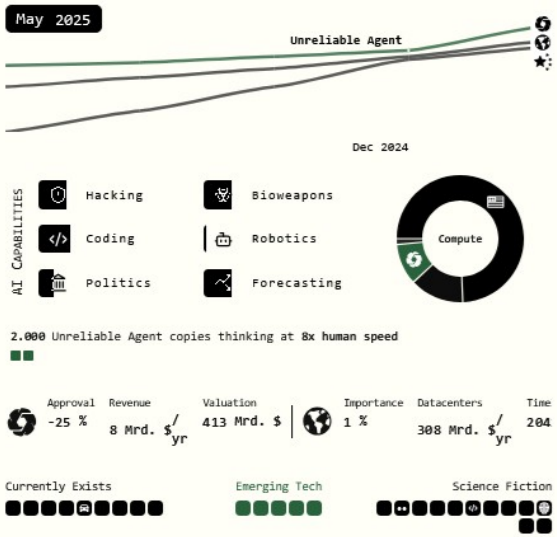
We wrote a scenario that represents our best guess about what that might look like. It's informed by trend extrapolations, wargames, expert feedback, experience at OpenAI, and previous forecasting successes.[2]

What is this? | How did we write it? | Why is it valuable? | Who are we?

Published April 3rd 2025 | PDF | Listen | Watch

Mid 2025: Stumbling Agents

The world sees its first glimpse of AI agents.



Isaac Asimov's 3(+1) Laws of Robotics (1942)

- 0th law:
  - A robot may not harm humanity,
    or, by inaction, allow humanity to come to harm.
- 1st law:
  - A robot may not injure a human being
    or, through inaction, allow a human being
    to come to harm.
- 2nd law:
  - A robot must obey the orders given it
    by human beings, except where such orders
    would conflict with the 1st law.
- 3rd law:
  - A robot must protect its own existence as long as such protection does not conflict with the 1st or 2nd law.

Today's Challenges (among others)

- Transparency
  - Does AI (and/or its training data) need to be open source?
- Algorithmic bias
  - How to measure & prove? Is there useful bias?
- Liability
  - If an AI does harm, who is legally in charge?
- Deceptive AI
  - What if an AI deceives humans to achieve its goals?
- Human dignity
  - Are there areas which should not be controlled by AI?

EC Ethics Guidelines

- AIs should be...
  - “lawful, complying with all applicable laws and regulations”
  - “ethical, ensuring adherence to ethical principles and values”
  - “robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm”

EC Ethics Guidelines

- Checklist
  - Human agency and oversight
  - Technical Robustness and safety
  - Privacy and data governance
  - Transparency
  - Diversity, non-discrimination and fairness
  - Societal and environmental well-being
  - Accountability

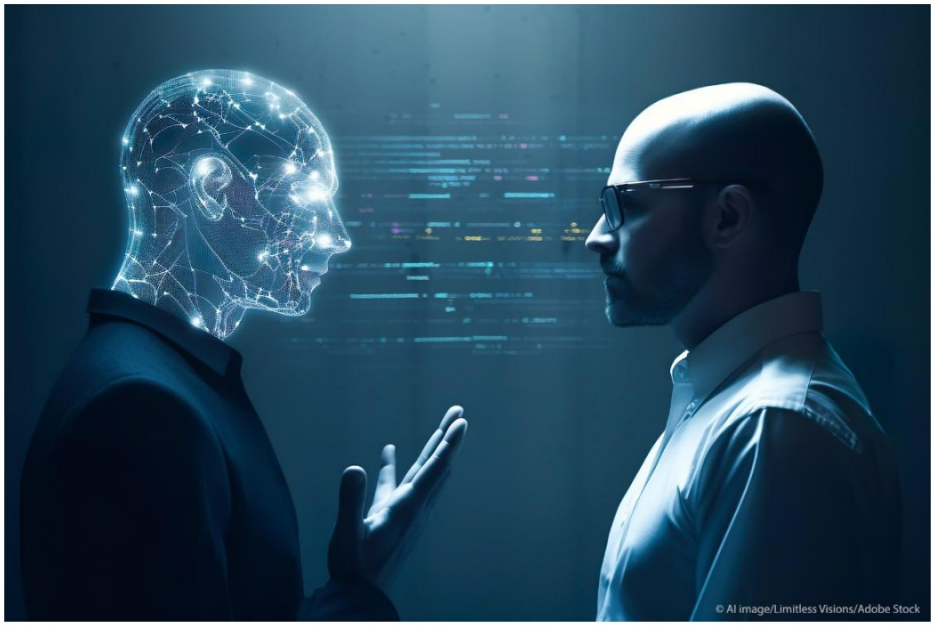
The EU AI Act

- Three categories
  - Unacceptable risk
  - High risk
  - Normal risk

EU AI Act: first regulation on artificial intelligence

Society Updated: 14-06-2023 - 14:06
Created: 08-06-2023 - 11:40

The use of artificial intelligence in the EU will be regulated by the AI Act, the world's first comprehensive AI law. Find out how it will protect you.



This illustration of artificial intelligence has in fact been generated by AI

The EU AI Act

- Unacceptable risk:
  - Cognitive behavioural manipulation of people or specific vulnerable groups (e.g. children)
  - Social scoring: classifying people based on behaviour, socio-economic status or personal characteristics
  - Real-time and remote biometric identification systems, such as facial recognition
- Are prohibited in the EU

The EU AI Act

- High risk:
  - AI systems used in products falling under the EU's product safety legislation (e.g., toys, aviation, cars, medical devices, ...)
  - AI systems in the following eight areas:
    - Biometric identification and categorisation of natural persons
    - Management and operation of critical infrastructure
    - Education and vocational training
    - Employment, worker management and access to self-employment
    - Access to and enjoyment of essential private services and public services and benefits
    - Law enforcement
    - Migration, asylum and border control management
    - Assistance in legal interpretation and application of the law
- Those undergo special assessment and audits

The EU AI Act

- Given those categories
  - Unacceptable Risk
  - High Risk
  - Acceptable Risk
- Where do you see ChatGPT and the like?

- AI Act introduces fourth category
  - General Purpose AI
  - Distinguishes GPAI which presents systemic risks vs. GPAI which doesn't
  - Systemic risk: Training compute >10^{25} FLOPs

What's that?

Joseph Weizenbaum's AI-free Areas (1976)

- AIs should not be used for…
  - A customer service representative
  - A therapist
  - A nursemaid for the elderly
  - A soldier
  - A judge
  - A police officer
- Where are we almost 50 years later?
- Do you agree with all of the above?

Thought Experiment: Universal Paperclips