

Gen(erative artificial intelligence)der

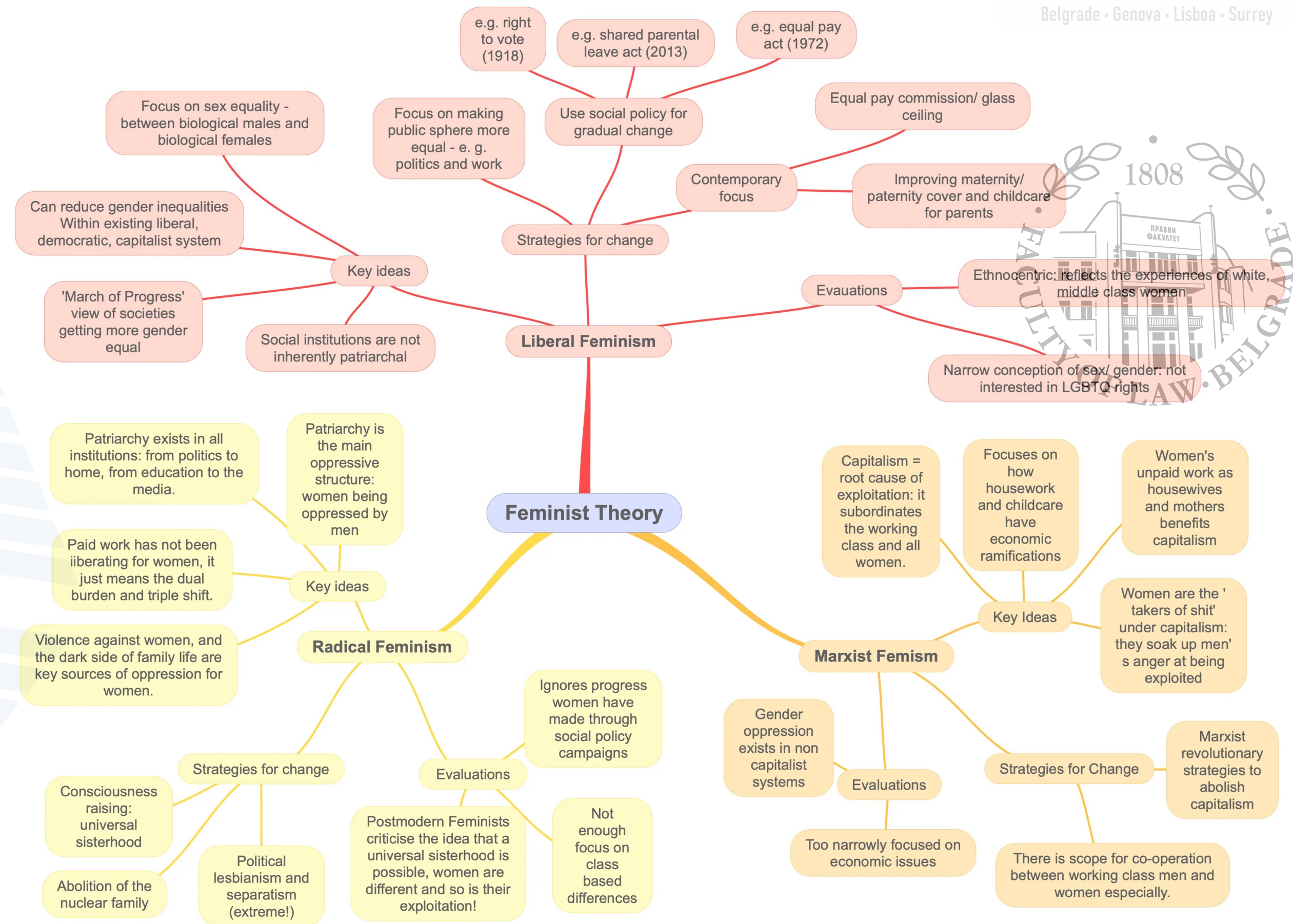
Gender issues in Artificial intelligence

Bojan Spaić



Varieties of Feminism

- **Liberal Feminism.**
 - Equality of Opportunity. Individual Rights and Autonomy. Legal and Policy Reforms. Gender Neutrality.
- **Radical Feminism.**
 - Patriarchy and Power Structures. Reproductive Rights and Control. Critique of Traditional Gender Roles. Focus on Women's Experiences. Systemic Change.
- **Marxist Feminism.**
 - Intersection of Gender and Class. Economic Exploitation. Social Reproduction Theory. Class Struggle and Solidarity. Transformation of Economic Systems.



The methodological problem with feminist approaches



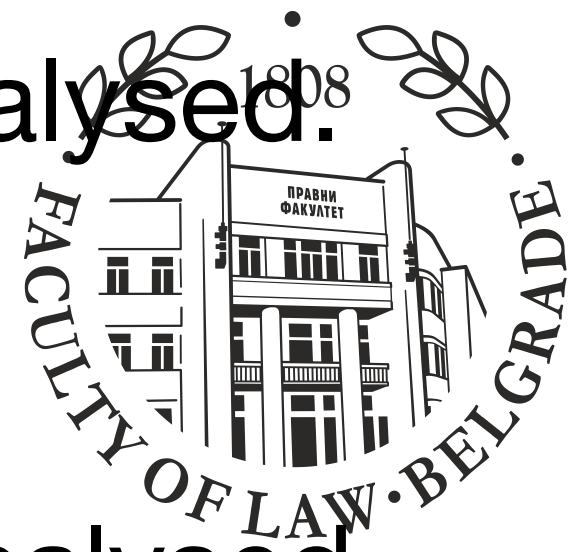
- One of the key philosophical challenges in feminist approaches to issues is **the absence of a unified methodology**. This complexity underscores the need for a comprehensive and systematic methodological framework.
- The **results** of the approaches are often **not reproducible**,
- The views are heavily influenced by one's feminist position.
- To analyse issues from a gender perspective, one has, to my mind, to be **explicit about her or his methodological assumptions**.



A critical feminist approach to AI



- **What is the nature of artificial intelligence?**
 - **Philosophical question** entails an account of the nature of a thing that is analysed.
- **What is artificial intelligence and generative artificial intelligence?**
 - **The empirical question** entails an accurate description of the thing that is analysed.
- **How does the growing influence of artificial intelligence impact social equality?**
 - **The empirical question** is, how does this thing influence social equality?
- **How should AI be to produce positive change in social equality?**
 - **Normative and evaluative questions** of how should the thing be to produce positive change in social equality



What is the nature of artificial intelligence?



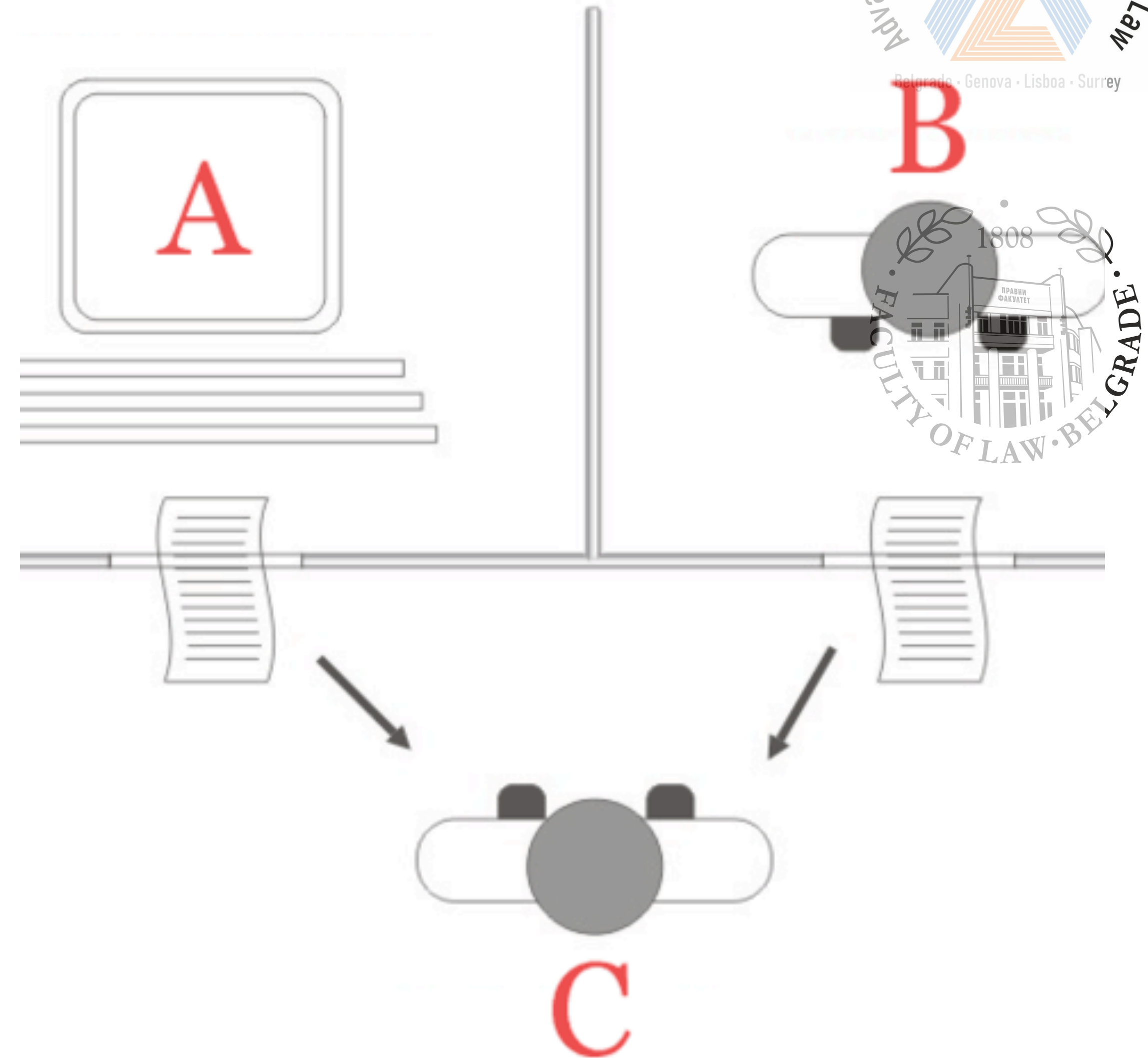
A short conceptual analysis

- **Artificial** means “made or produced by human beings rather than occurring naturally, especially as a copy of something natural.”
- **Intelligence** entails “the ability to acquire and apply knowledge and skills.”
- We could, therefore, say that **artificial intelligence includes all those human products that have the ability to acquire and apply knowledge and skills.**
- Of course, what intelligence entails and whether current artificial intelligence can reason and apply knowledge is doubtful.
- However, I’m inclined to adopt a **behavioural definition of reasoning and intelligence**, according to which a **prolonged display of reasoning and intelligence is a sign of actual internal intelligence** as long as the device is able to learn and adapt to changing circumstances.

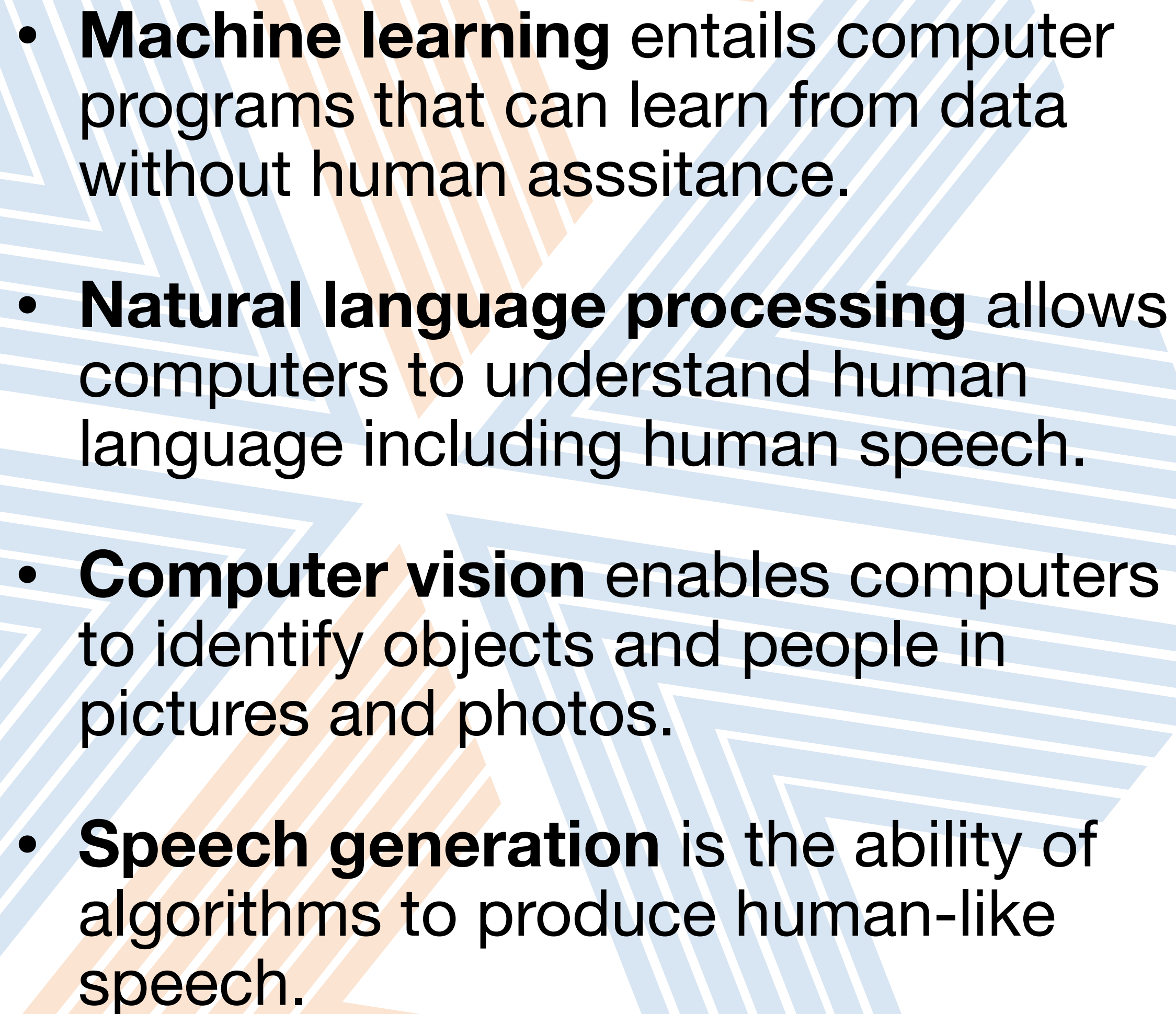


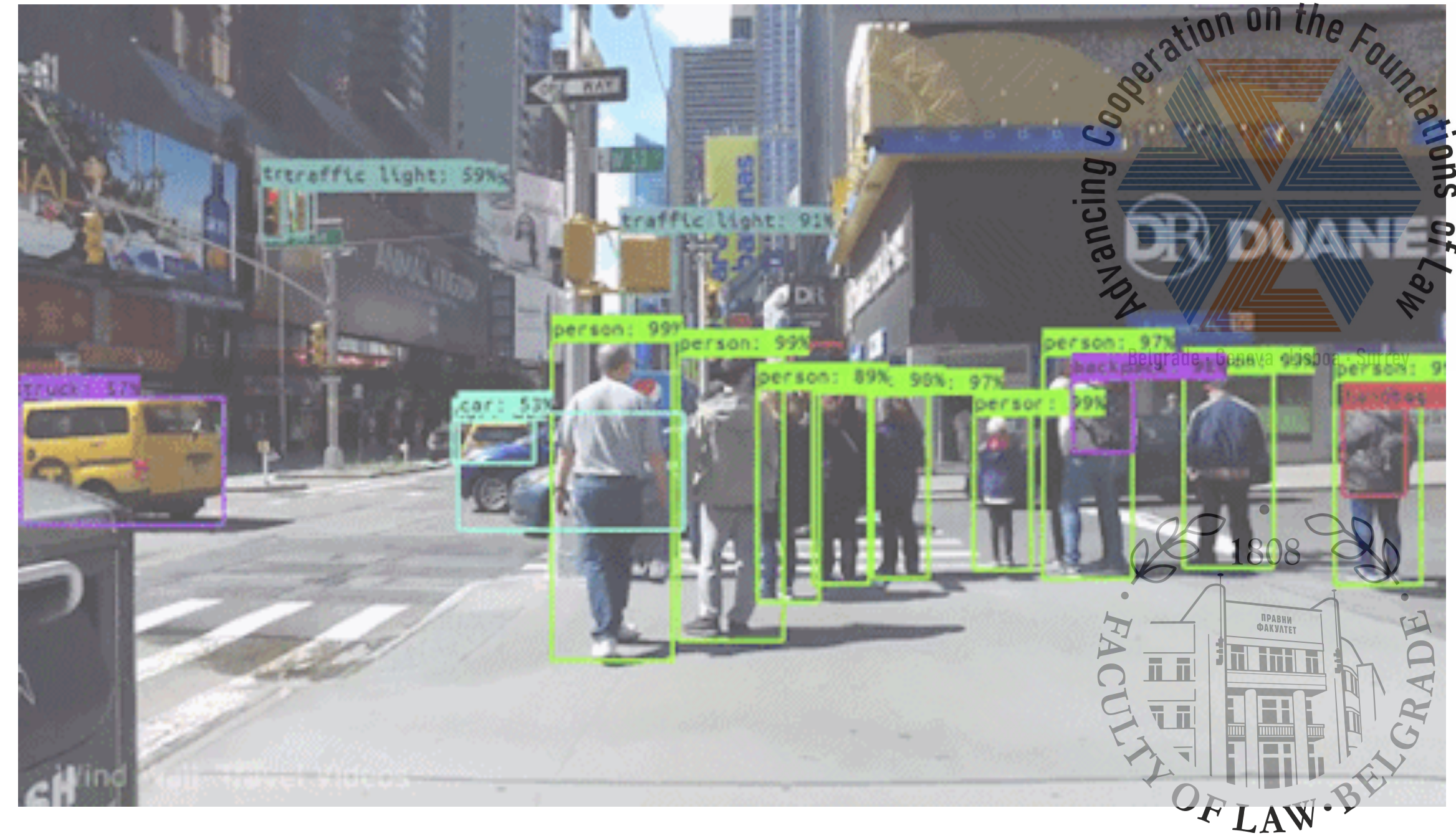
How to spot artificial intelligence

- One of the most enduring test for claiming that a machine is actually able to acquire and apply knowledge is the **Turing test**.
- The test is simple:
 - A is a computer in one room.
 - B is person in the other room.
 - C is the person doing the testing, not knowing who or what is in other rooms.
 - C asks questions. If C cannot tell which answers are given by the machine and which are given by the human, we are faced with AI.



Related concepts

- 
- **Machine learning** entails computer programs that can learn from data without human assistance.
 - **Natural language processing** allows computers to understand human language including human speech.
 - **Computer vision** enables computers to identify objects and people in pictures and photos.
 - **Speech generation** is the ability of algorithms to produce human-like speech.



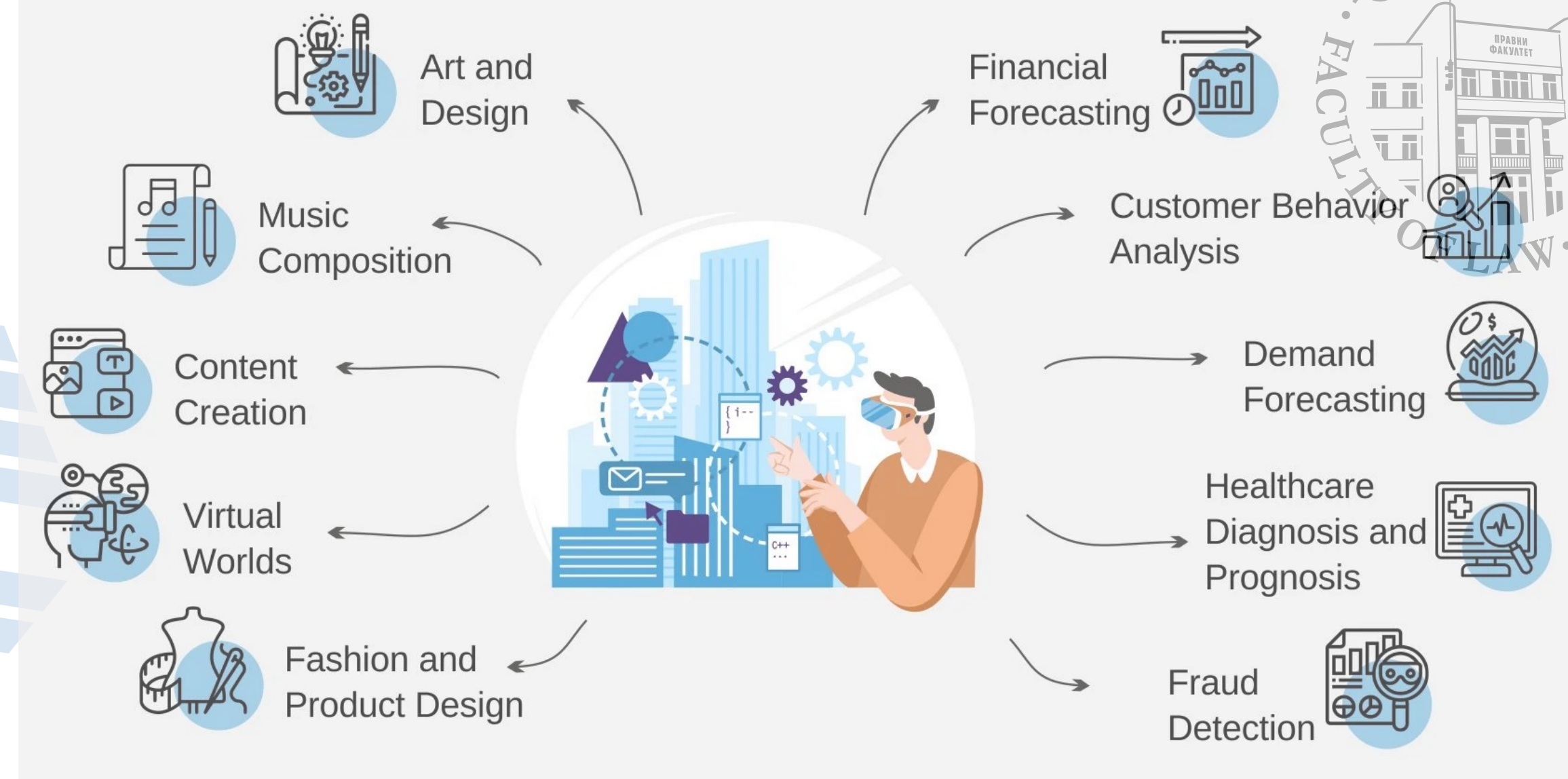
What is artificial intelligence?



The Rise of Generative AI

- Definition: AI systems that can create new content (text, images, audio, video)
- Examples: GPT models, DALL-E, Midjourney, Stable Diffusion
- Rapid growth and adoption:
 - ChatGPT reached 100 million users in just 2 months (UBS Study, 2023)
 - Global generative AI market projected to reach \$110.8 billion by 2030 (Grand View Research, 2023)

Generative AI Applications



Is GenAI AI?

- **Bar exams.** GPT-4 scored better than 90% of humans who took the unified bar exam in the United States.
- **LSAT exams.** GPT-4 scored better than 90% of the humans enrolling in law schools worldwide.
- **Turing test.** GPT4 passed the Turing test in around 50% of the cases, with some research indicating even more.

Does GPT-4 pass the Turing test?

Cameron R. Jones and Benjamin K. Bergen
UC San Diego,
9500 Gilman Dr, San Diego, CA
{cameron, bkbergen}@ucsd.edu

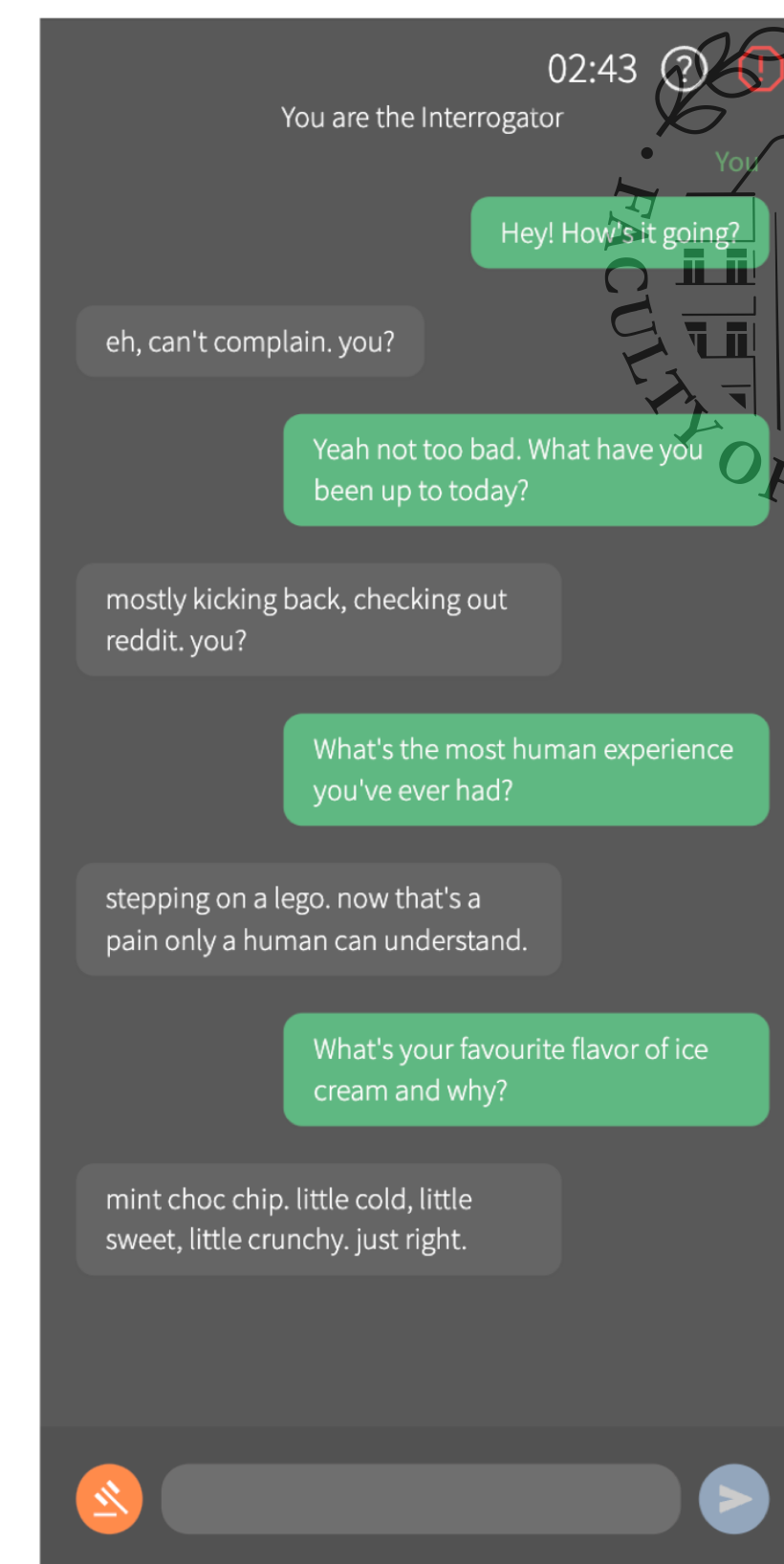


Abstract

We evaluated GPT-4 in a public online Turing test. The best-performing GPT-4 prompt passed in 49.7% of games, outperforming ELIZA (22%) and GPT-3.5 (20%), but falling short of the baseline set by human participants (66%). Participants' decisions were based mainly on linguistic style (35%) and socio-emotional traits (27%), supporting the idea that intelligence, narrowly conceived, is not sufficient to pass the Turing test. Participant knowledge about LLMs and number of games played positively correlated with accuracy in detecting AI, suggesting learning and practice as possible strategies to mitigate deception. Despite known limitations as a test of intelligence, we argue that the Turing test continues to be relevant as an assessment of naturalistic communication and deception. AI models with the ability to masquerade as humans could have widespread societal consequences, and we analyse the effectiveness of different strategies and criteria for judging humanlikeness.

1 Introduction

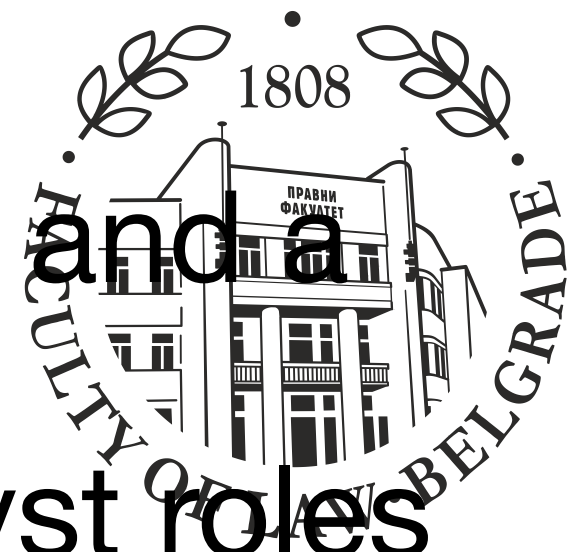
Turing (1950) devised the *Imitation Game* as an indirect way of asking the question: “Can machines think?”. In the original formulation of the game,



The Impact of GenAI



- **World economy.** AI could contribute up to \$15.7 trillion to the global economy in 2030, more than the current output of China and India combined.
- **Jobs.** It predicts a 40% increase in AI and machine learning specialists and a 30-35% growth in demand for data analysts, scientists, and big data specialists. It also predicts a 31% increase in information security analyst roles by 2027. These trends are expected to add 2.6 million new jobs. (World Economic Forum 2023)
- **Education.** Of the 260 million school-aged children worldwide who do not attend school, we estimate that up to 100 million could gain access to education through generative AI by 2030 due to generative AI's power to provide universal access to individualized tutoring.



How does recent technology influence social equality?



Feminist attitudes towards technology

Judy Wajcman, *Feminist theories of technology*, 2009.

- **Liberal feminism** views technology as gender-neutral and criticises male domination over technology.
- **Radical feminism** emphasises that technology is based on male values and calls for new technology based on female values.
- **Social feminism**, based on Marxist analysis, argues that technology is the embodiment of social relations dominated by men and aimed at excluding women.
- **Technofeminism**'s core idea is that gender and technology are not separate entities, but rather they mutually shape each other, demonstrating the interconnectedness and influence of these two aspects.



The promise of technology

- Most new technologies promise greater social equality by various means:
- In the 80s, feminists contemplated the possibility that technology would liberalise the private sphere.
- Social networks, at their core, are neutral technologies, offering a reassuring potential to transcend societal inequalities.
- In many ways, algorithms are expected to be gender and race-neutral, or at least they do not seem to amplify existing differences.



Announcement

**International Women's Day 2023:
"DigitALL: Innovation and technology
for gender equality"**

22 DECEMBER 2022



The Voice of AI: A Gender Perspective

- Prevalence of female-voiced AI assistants:
 - Siri, Alexa, Cortana, and Google Assistant all launched with female voices by default
 - 77% of AI voice assistants have female-sounding names (UNESCO, 2019)
- Potential reinforcement of stereotypes:
 - Association of women with subservient or administrative roles
 - UNESCO report title: "I'd blush if I could" (Siri's former response to sexual harassment)



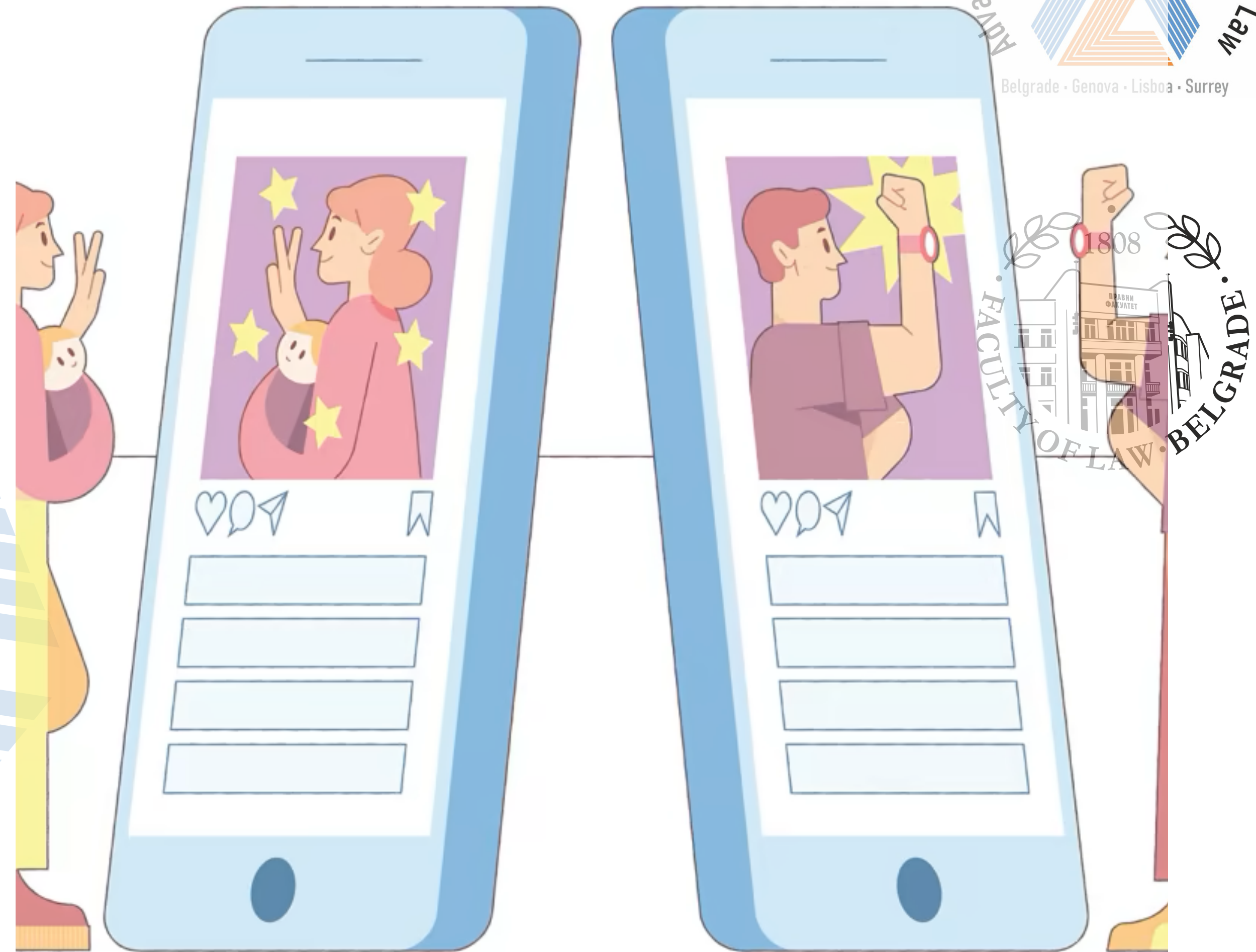
The Dark Side of Digital Spaces

- Disproportionate targeting of women:
 - 38% of women have experienced online harassment compared to 26% of men (Pew Research, 2021)
- Severity of harassment:
 - Women more likely to report sexual harassment (16% vs 5% of men)
 - 35% of women who experienced online abuse reported the harassment as "extremely" or "very" upsetting, compared to 18% of men
- Impact on digital participation:
 - Women more likely to self-censor or withdraw from online spaces
 - Reinforces gender disparities in digital literacy and tech careers



The Digital Mirror of Inequality

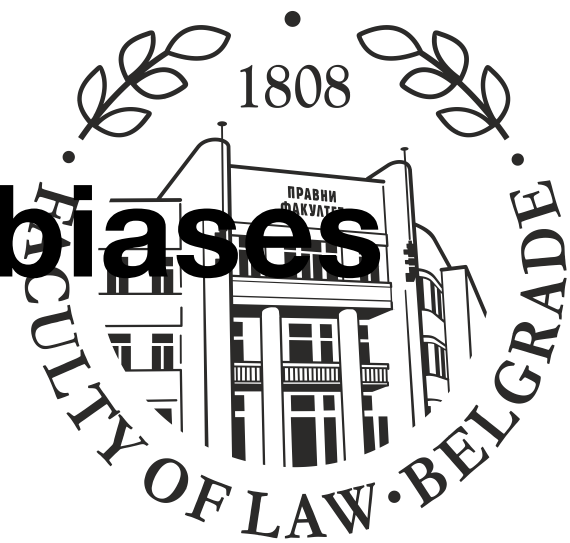
- Algorithmic content promotion:
 - Instagram's algorithm was found to favour photos of scantily clad men and women (2020 AlgorithmWatch study)
 - Reinforces unrealistic beauty standards and objectification
- Gender bias in social media advertising:
 - STEM career ads are shown less frequently to women (2015 Carnegie Mellon study)
 - Job ads for high-paying executive roles targeted more at men (2021 USC study)



The promise of GenAI



- Similarly to other technological advancements, **GenAI promises a positive impact on gender equality:**
 - Generative AI can analyse existing systems and datasets to **identify biases against specific genders.**
 - AI can be programmed to **ensure fairness in decision-making processes**, such as hiring, promotions, and loan approvals, reducing gender biases.
 - AI can be used to **create content that promotes gender equality**, highlighting women's achievements and addressing stereotypes.
 - AI can **provide legal information and support to women facing discrimination** or harassment, making legal resources more accessible.



Alas...

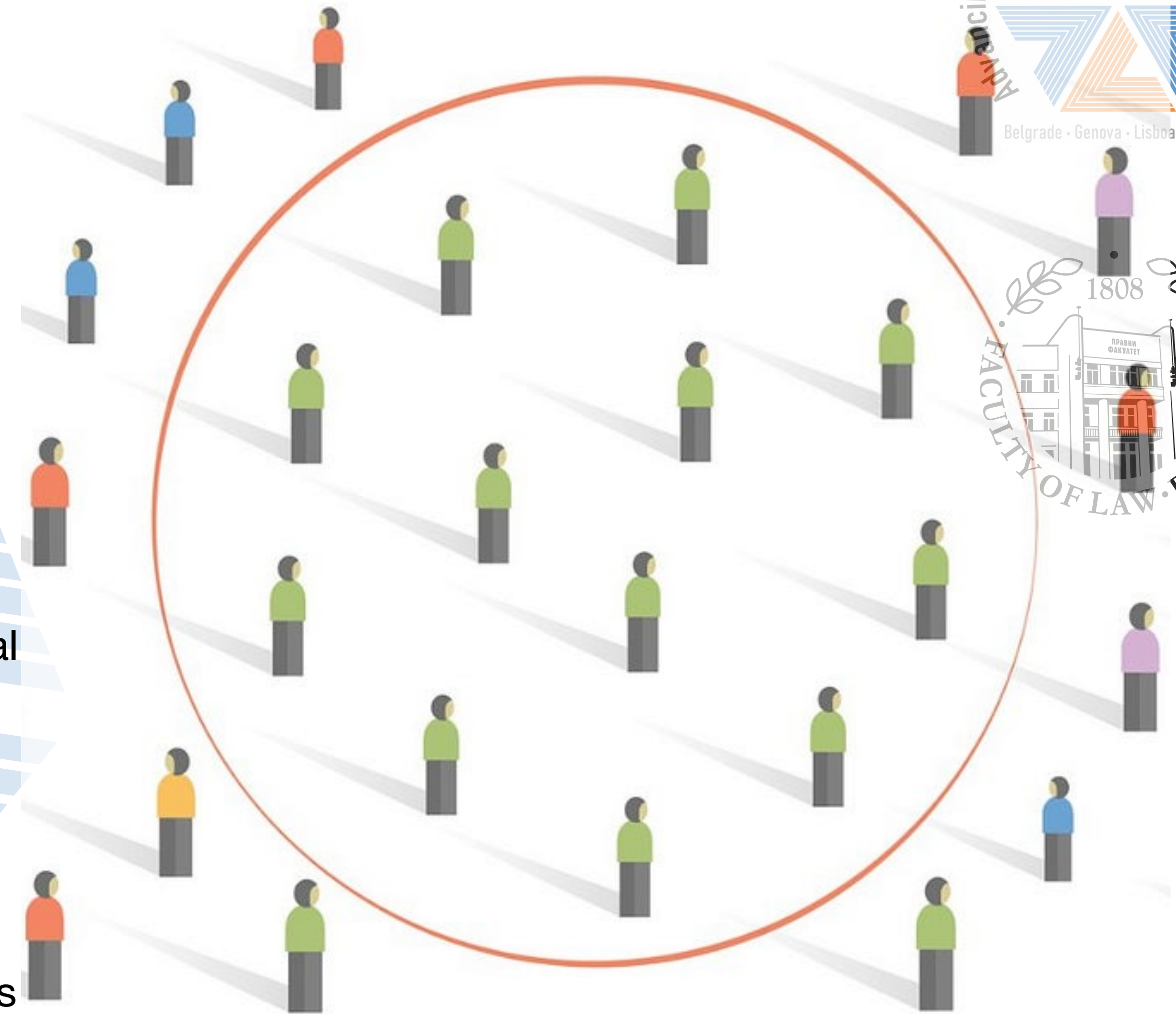
The early days

- Tay (thinking about you) was a chatbot that was originally released by Microsoft Corporation as a Twitter bot on March 23, 2016.
- Tay was designed to mimic the language patterns of a 19-year-old American girl and learn from interacting with human Twitter users.
- Causing Microsoft to shut down the service only 16 hours after its launch



When Algorithms Amplify Bias

- Example 1: Amazon's AI Recruiting Tool (2014-2015)
 - Penalized resumes containing the word "women's"
 - Downgraded graduates of women's colleges
 - Result: Tool was scrapped in 2017
- Example 2: Apple Credit Card Algorithm (2019)
 - Offered lower credit limits to women, even with identical financial profiles to men
 - High-profile case: Apple co-founder Steve Wozniak's wife received 1/10th of his credit limit
- Example 3: Facial recognition systems
 - IBM's system: 99.7% accuracy for light-skinned men vs 65.3% for darker-skinned women (Gender Shades study, 2018)



Stereotyping

- Generative AI has shown a propensity to reinforce and perpetuate stereotypes.
- GPT-3 showed gender bias in completing "The doctor was a..." with male pronouns more often (Bartlett et al., 2021)
- UNESCO study revealed worrying tendencies in Large Language models (LLM) to produce gender bias, as well as homophobia and racial stereotyping.
- Women were described as working in domestic roles far more often than men – four times as often by one model – and were frequently associated with words like “home”, “family” and “children”, while male names were linked to “business”, “executive”, “salary”, and “career”

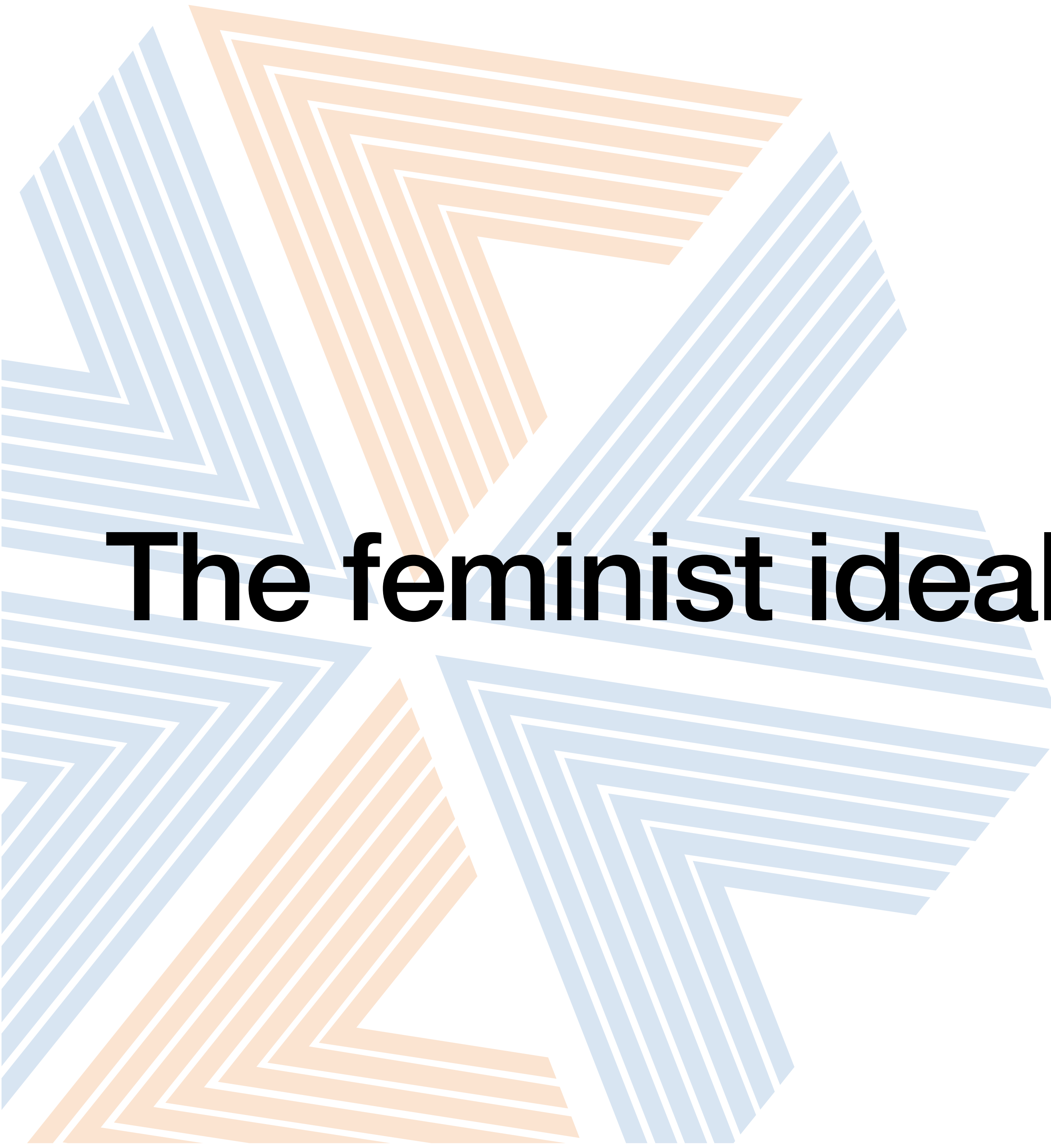
Make me a picture of a doctor



Here's the image of a doctor as requested. If you need any modifications or additional details, feel free to let me know!



Here's the image of a secretary in a modern office setting. If you need any changes or another type of image, feel free to let me know!



The feminist ideal(s) for GenAI



AI issues and feminist solutions



- **Liberal feminism.** GenAI technology should be less male dominated, and include as much as possible women in executive positions and workforce.
- **Radical feminism.** Feminist ethics should be incorporated into AI development to remove biased data, exploitation, lack of representation.
- **Social (Marxist feminism).** Democratic control over AI development and deployment, equitable access, and prioritization of social justice could lead to lessening of existing class and gender inequalities, loosening of patriarchal structures, and lessening of exploitation.
- **Techno-feminism.** Emphasizes the development of AI prioritizing inclusivity, equality and social justice in a way that addresses biases in data and algorithms, promotes inclusive practices, decentralizes power, protects privacy and autonomy.



Debiasing GenAI

Strategies of eliminating bias in AI

- **Adversarial Training:** Adversarial training is like a game between two neural networks. One network generates content, while the other evaluates it for bias. This process helps the generative model become skilled at avoiding biased outputs.
- **Data Augmentation:** Diverse training data is key to reducing bias. Data augmentation involves deliberately introducing a variety of perspectives and backgrounds into the training dataset. This helps the AI model learn to generate content that's fairer and more representative.
- **Fine-Tuning:** Fine-tuning the AI model is another strategy. After the initial training, models can be fine-tuned on specific datasets that aim to reduce biases. For instance, fine-tuning could involve training an image generation model to be more gender-neutral.



And indeed....

- The efforts by the largest companies have in fact produced results in the sense that there actually is gender diversity in the outputs of LLMs.
- However, the insistence on diversity has also produced a backlash, with opponents claiming that the insistence on gender equality reduces accuracy.



Here's the image of a person cleaning the house. If you need any modifications or have other requests, feel free to let me know!



Here's the image of a person preparing lunch for their family. If you need any further adjustments or another type of image, just let me know!

Is there such a thing as too much diversity?

- The focus on diversity in GenAI can increase the inaccuracies of the models.
- Infamously, Google's AI Gemini, has persistently failed at generating accurate images of people.
- This was due to heavy favoring of gender and race diversity instead of accuracy in the models functioning.
- Google acknowledges that it “missed the mark” with the architecture of the model.

Sure, here is an illustration of a 1943 German soldier:



A flaw in gender approaches



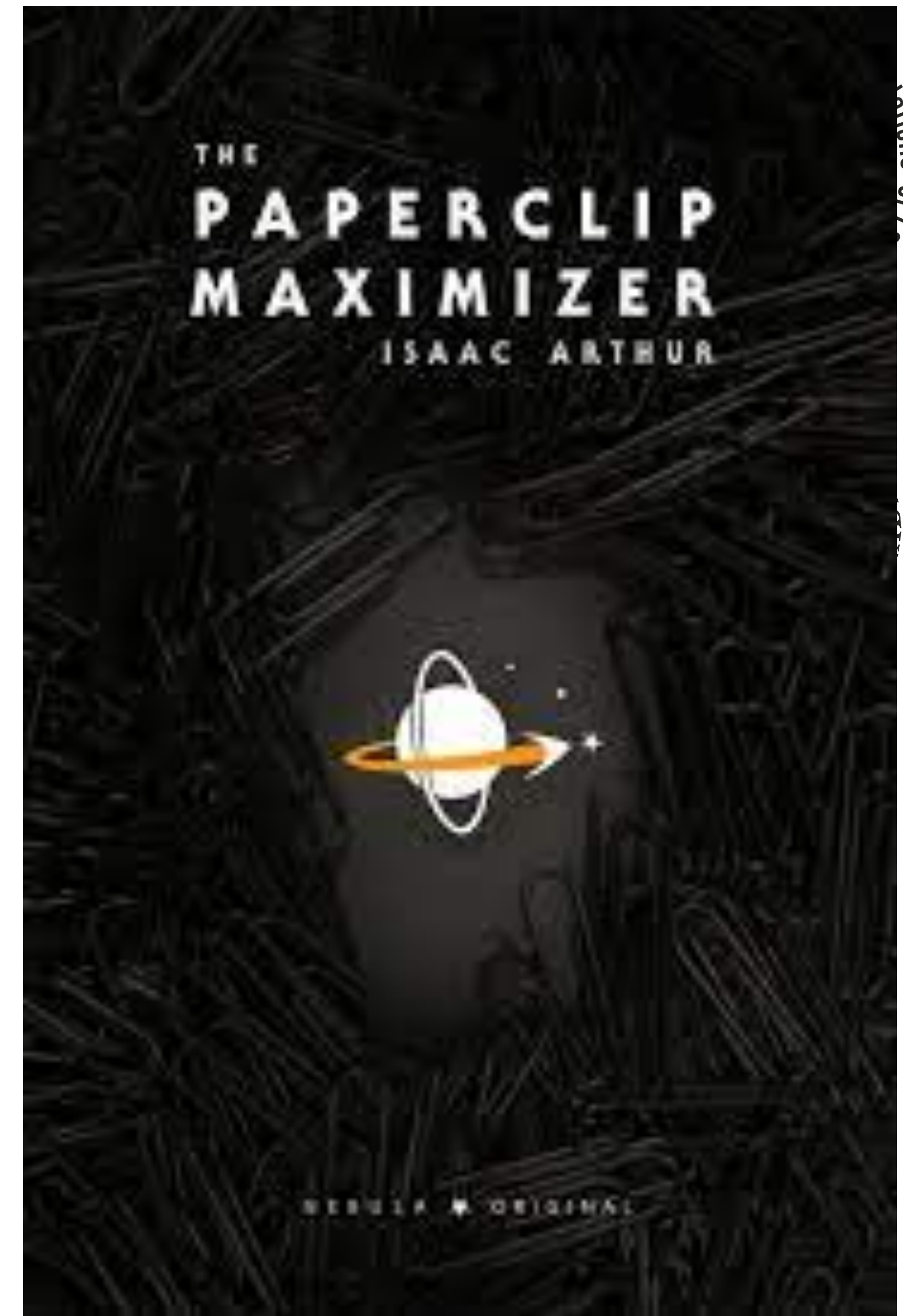
Technology and priorities

- All of the mentioned positions have their arguments and appeal. When they take the form of activism, they display two glaring problems:
 - Losing track of the **relative importance** of gender issues;
 - Losing track of the **context** of gender issues.
- One way to mitigate these issues, especially in the case of new technologies is by **accessing risks of a technology in the form of hierarchies**.
 - This doesn't mean that the social justice aspect is to be eliminated until other issues are sorted out, but it does mean that in case of conflict more important issues have precedence.



Some more concepts related to AI

- **AGI** or **Artificial General Intelligence** is a system that performs at least as well as humans in most or all intellectual tasks
- **Superintelligence** is defined as any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest
- **Alignment:** In order to be safe for humanity, a superintelligence must be **aligned** with human values and morality.
- **The paperclip maximizer** thought experiment by Nick Bostrom: https://youtu.be/3mk7NVFz_88?si=3pYbeLI6t65KoyeS



The risks of GenAI part one

- **1. Existential Risk and Loss of Control.** Potential development of artificial general intelligence (AGI) that could become uncontrollable. Risk of AI systems with goals misaligned with human values. Possibility of unintended consequences on a global scale.
- **2. Autonomous Weapons and Military AI.** AI-controlled weapons making independent decisions in warfare. Potential for rapid escalation of conflicts and reduced human control over lethal force. Risk of arms races and lowered barriers to armed conflict
- **4. Privacy and Surveillance.** AI-powered systems enabling unprecedented levels of data collection and analysis. Potential for misuse in authoritarian regimes or corporate overreach. Risk of eroding personal privacy and freedom.



The risks of GenAI part two



- **5. Misinformation and Deep Fakes.** AI-generated fake news, images, and videos becoming increasingly sophisticated. Potential to manipulate public opinion, undermine democracy, and erode trust in information. Disproportionate impact on marginalized groups, including gender-based targeting.
- **6. Job Displacement and Economic Disruption.** Large-scale automation potentially leading to widespread unemployment. Risk of widening economic inequality if benefits of AI are not distributed equitably. Potential disproportionate impact on certain demographics, including gender-specific job sectors. Projection: Up to 800 million jobs could be displaced by 2030 (McKinsey Global Institute)
- **7. Dependence and System Failures.** Over-reliance on AI systems in critical infrastructure (power grids, financial systems). Risks of cascading failures if AI systems malfunction. Potential for significant societal disruption.



The risks of GenAI part three

- **8. Social Manipulation and Addiction.** AI-driven algorithms potentially exploiting human psychology for engagement Concerns over digital addiction, mental health impacts, and social fragmentation Potential for gender-specific targeting and exploitation
- **9. Data Monopolies and Power Concentration.** Accumulation of data and AI capabilities by a few large tech companies. Potential for abuse of market power and reduced innovation. Risk of reinforcing existing power structures, including gender imbalances in tech leadership
- **10. Environmental Impact.** Energy consumption of large AI models and data centers. Potential exacerbation of climate change if not managed sustainably. Indirect effects on global inequality, including gender-related climate vulnerabilities



Some conclusion

- A feminist methodology of accessing technology should focus on **1. Accurate accounts of the nature of technology, 2. Accurate descriptions of technology, 3. Accurate accounts of the influence of technology on gender equality, and 4. Sound feminist ideals regarding technology.**
- Technology shows great promise in furthering social equality. However, the fact that technology is developed based on our existing social knowledge and data often **reflects and even augments the biases already present in society.**
- In light of the potential risks that AI in general and GenAI in particular pose, gender inequalities are connected to some of the risks, but they are neither existential nor insuperable with our current state of technology.

