

## SUMMARY

### Societal scale risks

#### Misuse of AI:

- AI could enable criminal or terrorist activities
- Mass manipulation
- Pervasive surveillance



[3]

#### Unintentional consequences:

- Lose control of AI
- AI is making progress in hacking, social manipulation, deception
- Copy their algorithms across global server networks
- Control the computer systems behind our communication



[4]

## SUMMARY

### Path forward

#### Reorienting Technical R&D:

- Dedicate at least 1/3 of their AI R&D
- Pursue and research breakthroughs in AI safety and ethics

#### Governance Measures:

- Comprehensive insight into AI
- Analyze and test systems before launch
- International agreements and partnership

## EVALUATION

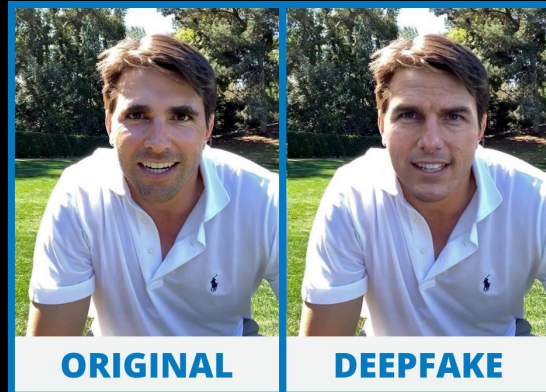
- While powerful, AIs aren't likely to become fully autonomous
- Current autonomous AIs aren't powerful
- The international accord idea is nice for a perfect world
  - We aren't in a perfect world
- Governments won't add laws since they benefit from AI progress
- Proposed regulations weren't fleshed out in the paper
- Doesn't consider full company impact

## ETHICAL CONCERNS

- Many concerns raised were about the misuse of AI
  - By people: autonomous warfare/surveillance, misinformation/deepfakes
  - By itself: copy itself across servers like worms, use weaponry against humans
- Proposed solution:
  - Heavy regulation of frontier systems
    - 1/3 rule for safety/ethics research
      - Would this stifle innovation, especially in smaller companies or countries?
    - Balance between safety and innovation
      - Is it ethical to halt research on tech that can improve many lives because of potential bad outcomes?
  - Worldwide agreements and partnerships
    - Seen in past that countries can circumvent or ignore agreements for their own gain
    - Can we stop bad actors from developing AI?

## OUR OWN THOUGHTS

- Crime and Terrorist Activities
- Security Vulnerabilities
- AI is a Tool
- Look to the Creators of that AI
- Speculation
- Regulations



[5]

## IMPACT TO SOCIETY

- Benefits
  - AI can perform better than us on many computationally heavy tasks
    - Opens the doors for potential disease cures, environmental management, etc.
- Concerns
  - Authors focus mostly on potential future problems
    - AI warfare, surveillance, and terrorist use
  - They focus little on problems facing us today
    - Misinformation, hallucinations, and deepfakes
  - We should focus more on addressing the problems impacting us now
    - We can do something about our current AI problems

## QUESTIONS

- Should companies devote 1/3 of their budget/resources towards safety/ethics research?
- At what point do we draw the line with AI between risk and progress? Should we stop the development of AI at some point?
- Is it realistic to have worldwide agreements and partnerships (like with the NPT)?
- Are we more concerned about AI taking over or people misusing AI?
- Do we think AI is going to take over any time soon?
- What other systems or laws could we put in place to help regulate/restrict the development of AI?

## REFERENCES

- [1] Bengio, Yoshua, et al. "Managing ai risks in an era of rapid progress." arXiv preprint arXiv:2310.17688 (2023).
- [2] Karjian, Ron. "The History of Artificial Intelligence: Complete Ai Timeline." *Enterprise AI*, TechTarget, 16 Aug. 2023, [www.techtarget.com/searchenterpriseai/tip/The-history-of-artificial-intelligence-Complete-AI-timeline](http://www.techtarget.com/searchenterpriseai/tip/The-history-of-artificial-intelligence-Complete-AI-timeline).
- [3] Belcic, Ivan. *What Is Hacking?*, Avast, 23 Feb. 2023, [www.avast.com/c-hacker](http://www.avast.com/c-hacker).
- [4] Wiki, Contributors to Villains. "Skynet." *Villains Wiki*, Fandom, Inc., [villains.fandom.com/wiki/Skynet](http://villains.fandom.com/wiki/Skynet). Accessed 1 May 2024.
- [5] -"Are Deepfakes All Evil When They Can Be Used for Good? " *Are Deepfakes All Evil When They Can Be Used for Good? - Video Personalization | Personalized Video Email Marketing | Maverick Videos*, 2024, [www.trymaverick.com/blog-posts/are-deep-fakes-all-evil-when-can-they-be-used-for-good](http://www.trymaverick.com/blog-posts/are-deep-fakes-all-evil-when-can-they-be-used-for-good).