

## Responsible AI Assignment 1 Ideas

Below is a list with suggested ideas for Assignment 1. You can choose one of these ideas or define your own idea.

1. Responsibility assessment for Large Language Models.
2. Transparency assessment for Large Language Models.
3. Safety, security, and privacy assessment for Large Language Models. The focus could be here also just one aspect instead of all of them.
4. Trustworthiness assessment for Large Language Models.
5. Trustworthiness of generative models applied in critical domains like cyber security, defense, or law enforcement.
6. Explainability of generative models applied in critical domains like cyber security, defense, or law enforcement.
7. Safety and security of generative models applied in critical domains like cyber security, defense, or law enforcement.
8. Security and privacy of generative models applied in critical domains like cyber security, defense, or law enforcement.
9. Fairness of generative models applied in critical domains like cyber security, defense, or law enforcement.
10. Multimodal generative models applied in critical domains like cyber security, defense, or law enforcement.
11. Trustworthiness of generative models.
12. Deepfakes in time of conflict or war.
13. Deepfakes detection and attribution.
14. Social media bots in time of conflict or war.
15. Generative models for cyber warfare or hybrid warfare gaming.
16. The use of AI in building security attacks.
17. The use of AI in building privacy attacks.
18. Autonomy in time of conflict/war for Unmanned Aerial Vehicles.
19. Decision making for autonomous cars.
20. Robotization (e.g., bias, artificial error).



21. AI-soldier of the future and its target selection (e.g., profiling, facial recognition).
22. Predictive policing.
23. AI robot lawyer.
24. Brain Computer Interfaces.
25. Multiverse (e.g., access, participation, conflict).
26. Social credit system (e.g., in China).
27. Climate modelling and simulation.
28. The Singularity.
29. Artificial General Intelligence (e.g., self-awareness, use in time of conflict/war etc.).
30. Artificial Super Intelligence decision making.
31. Artificial Super Intelligence used in / creating conflict.