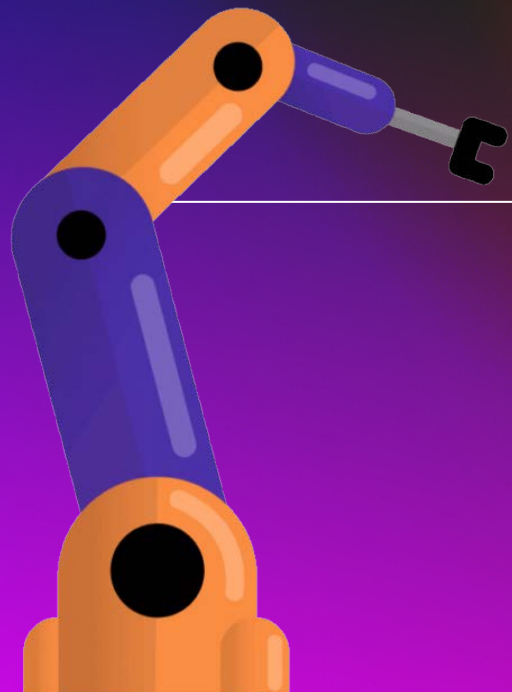




# BUILDING ETHICAL AI SYSTEMS

---



Victor van der Mark, Yannick Toonen & Rico de Bokx

PARIS

# **MICROSOFT'S RACIST ROBOT**

AI "CHATBOT" TAKEN OFFLINE AS TWEETS TURN OFF-COLOUR

BUSINESS

FRANCE24.COM





# Question for you!

**Are you scared or afraid of AI?**





















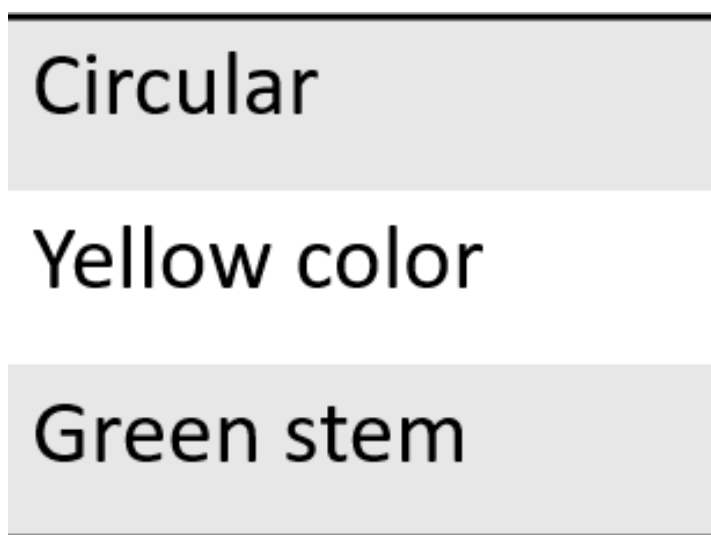




# AI 1: Supervised learning



**Labeled dataset**



**Criteria**



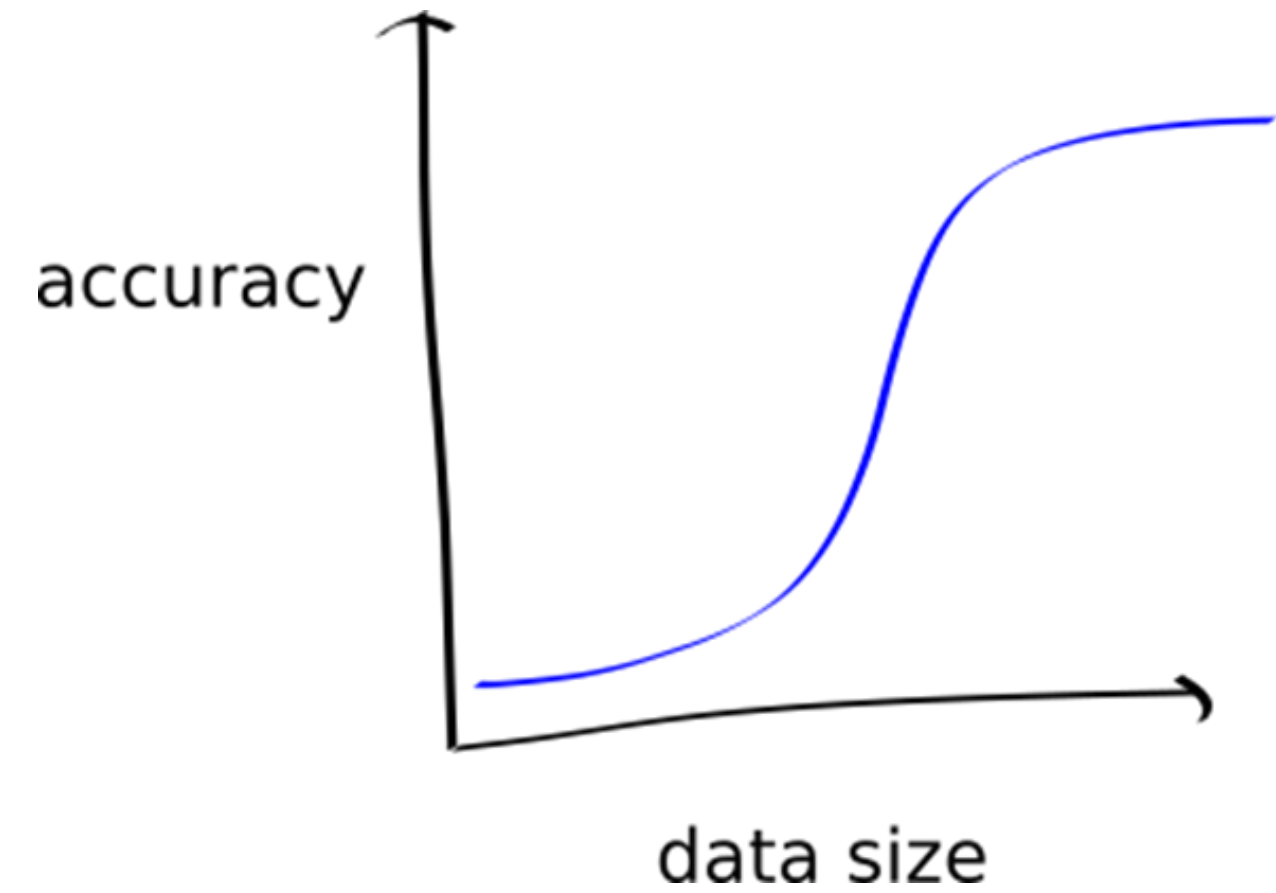
**Test data set**



**Guess**

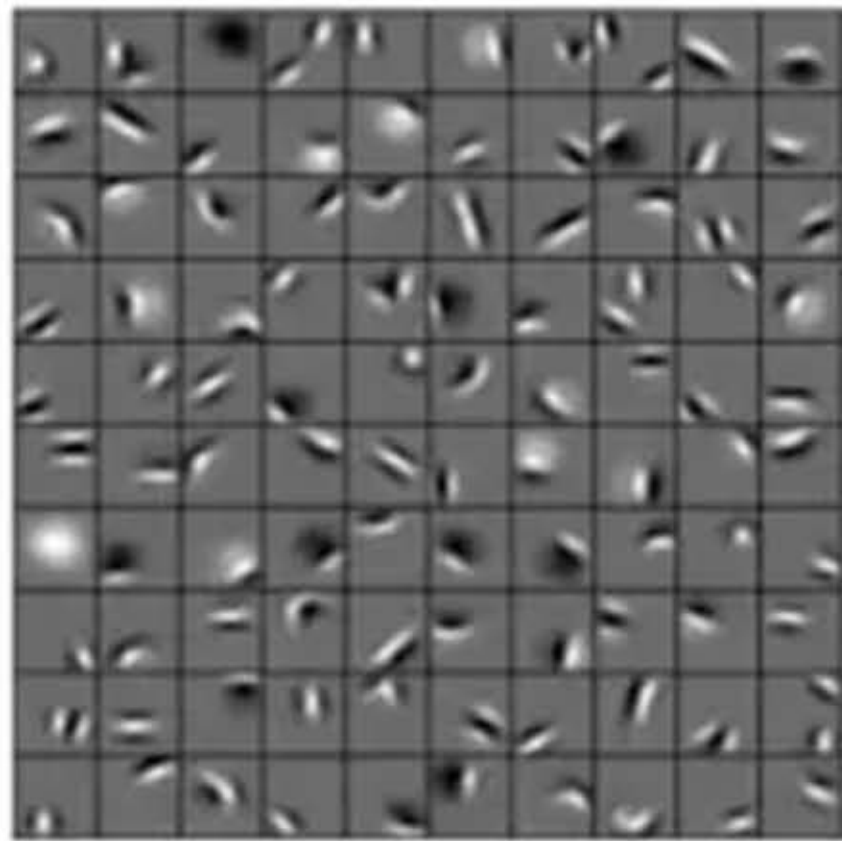


# Supervised learning & data





# AI 2: Unsupervised learning



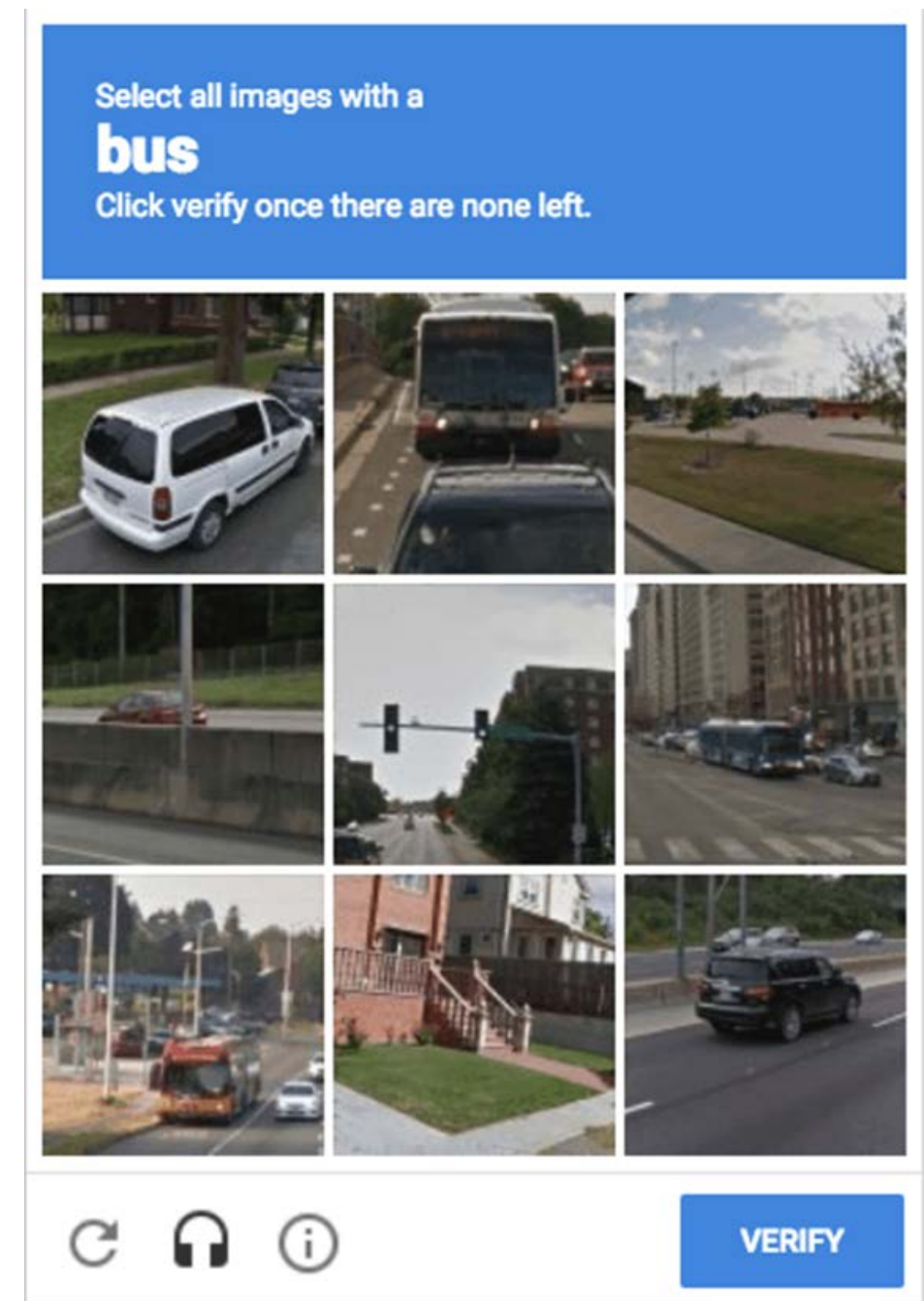
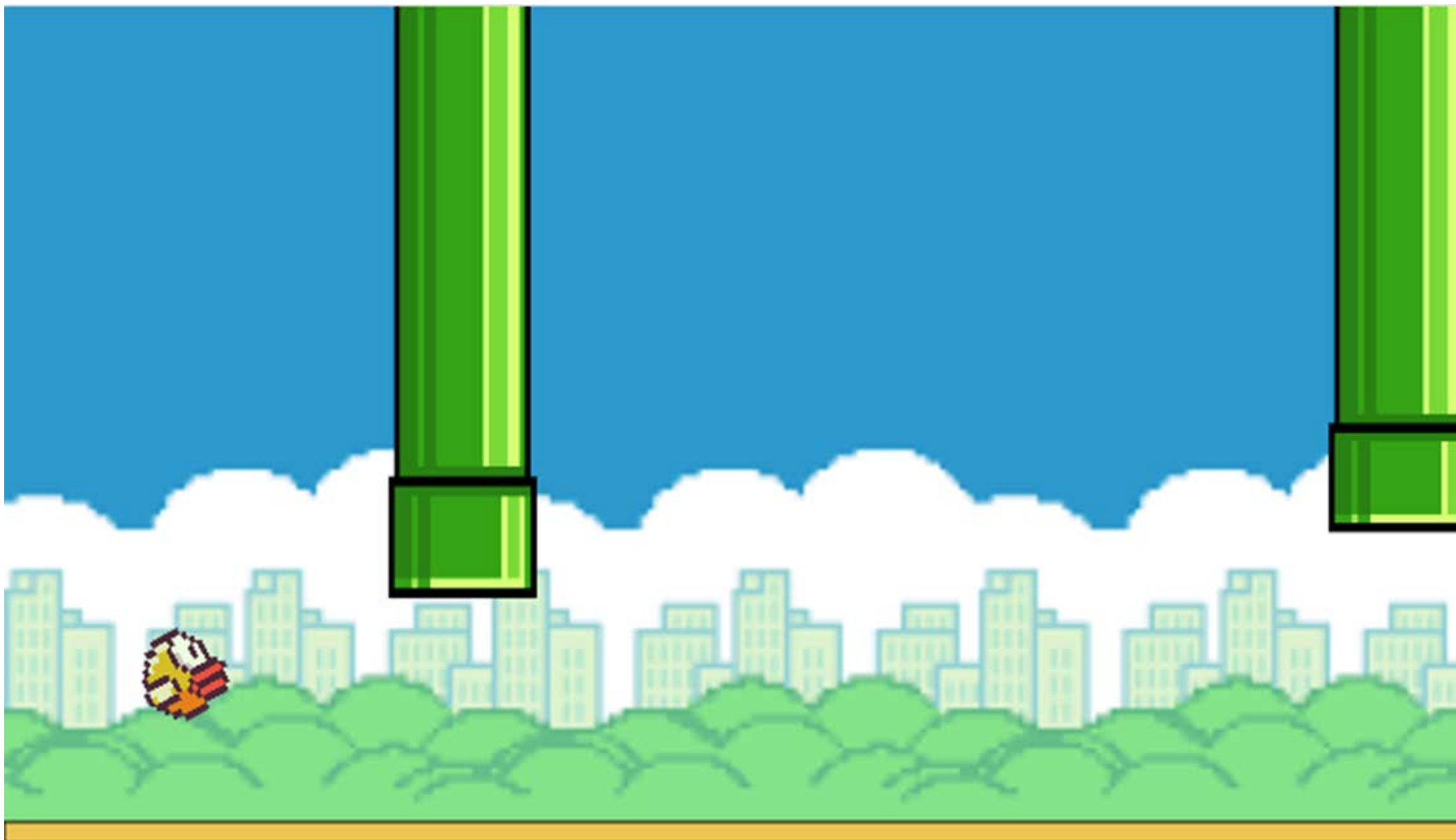


# AI 3: Semi-supervised learning





# AI 4: Reinforcement learning





**WHEN DO YOU THINK THE FIRST AI  
SYSTEM WAS DEVELOPED?**



# History of AI



**1959**

Arthur Lee Samuel chess machine

**2000-2010**

Various AI breakthroughs

**2020-present**

First widely used AI chatbot

**1980**

First commercial AI system

**2010-present**

Introduction of various consumer usable AI programs



# WHY IS AI GETTING POPULAR NOW?





# General types of AI

Reactive Machines	Limited Memory	Theory of Mind AI	Self-aware AI
<ul style="list-style-type: none"><li>• Simplest form</li><li>• Not able to learn</li><li>• Analysing the current state</li></ul>	<ul style="list-style-type: none"><li>• Ability to retain data</li><li>• Improving over time</li></ul>	<ul style="list-style-type: none"><li>• Aims to create machines that people understand</li><li>• Can recognize thoughts and emotions</li></ul>	<ul style="list-style-type: none"><li>• Deep understanding of its own existence</li><li>• Can reflect</li></ul>
Chess-playing AI	Self driving cars	Virtual Chat bots	-

*More intelligent*





# Steps for building an AI system





# Ethical Applications

---

**Ethically sensitive applications**

**Mitigation and solutions**

**Trust in creators**





# Facial Recognition

---

**Privacy Concerns**

**Biases**

**Consequences**







# Autonomous Vehicles

---

**Fast moral decisions**

**Accountability & Responsibility**



# Hiring & Recruitment

Discrimination

Unfair

Unrepresentative





# Mitigating Concerns

## Facial Recognition

---

Regulations &  
Guidelines

Audits

## Autonomous Vehicles

---

Industry Standards

Legislation

## Hiring & Recruitment

---

Transparency

Educating Recruiters



# Trust in Creators

---

Track Record

Continuous Improvement

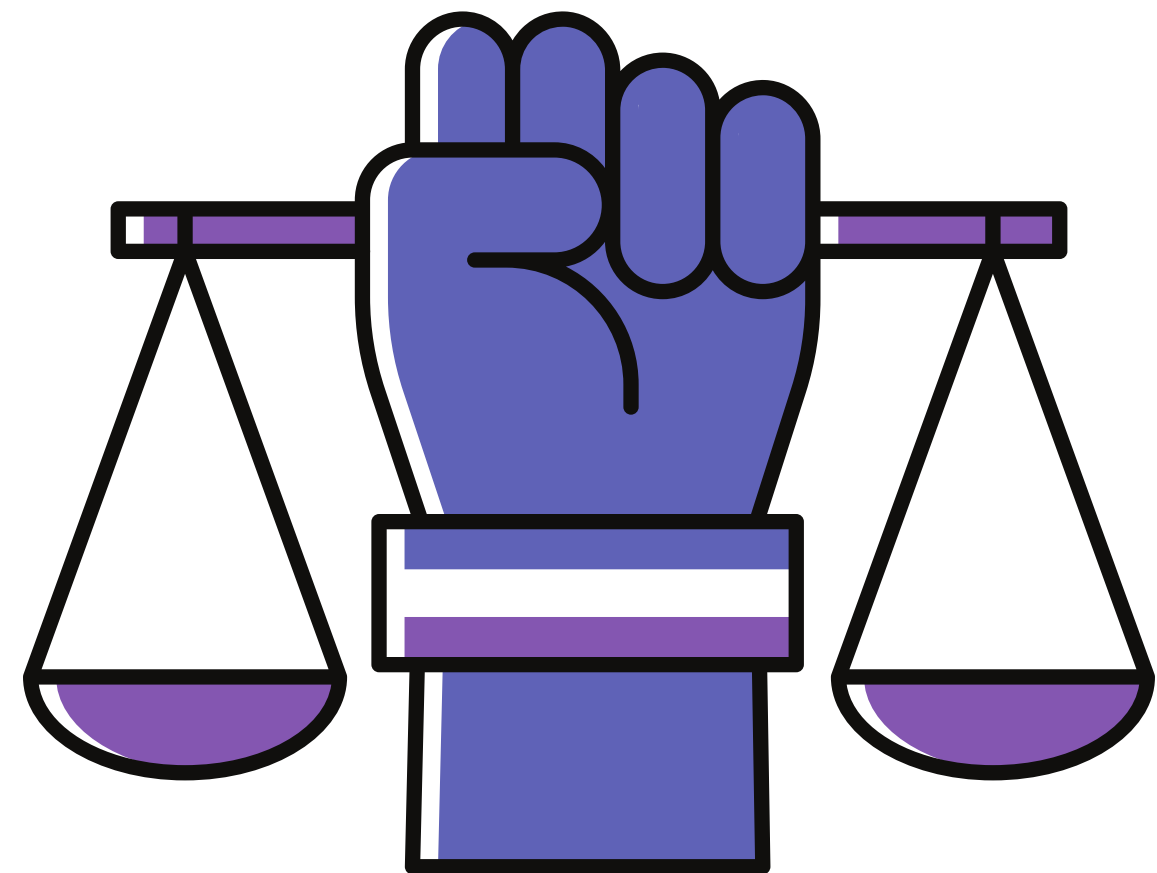
Red Team Assessments

Commitment to Collaborations





# Discussions





# ChatGPT needs to be implemented in school curriculums

---

**Agree**

**Disagree**



# Should AI make life-or-death decisions?

---

**Yes**

**No**



# Who is responsible for a crash of an AI controlled vehicle?

---

**Driver/owner**

**Manufacturer**



# AI should automate jobs even if it results in job loss for humans

---

**Agree**

**Disagree**



# AI systems should be given legal rights like humans and animals

---

**Agree**

**Disagree**



**It is ethical to create AI caretakers for elderly  
even if it decreases human interaction**

---

**Agree**

**Disagree**



**There need to be limitations on the development  
of AI that can surpass human intelligence**

---

**Agree**

**Disagree**



# I feel as safe in an autonomous car as a human-driven car

---

**Agree**

**Disagree**



# AI-generated art should be eligible for copyright protection

---

**Agree**

**Disagree**



# Should AI assist in recruitment?

---

**Yes**

**No**



# Are you (still) scared of AI?

---

**Yes**

**No**