

The Ethics of AI

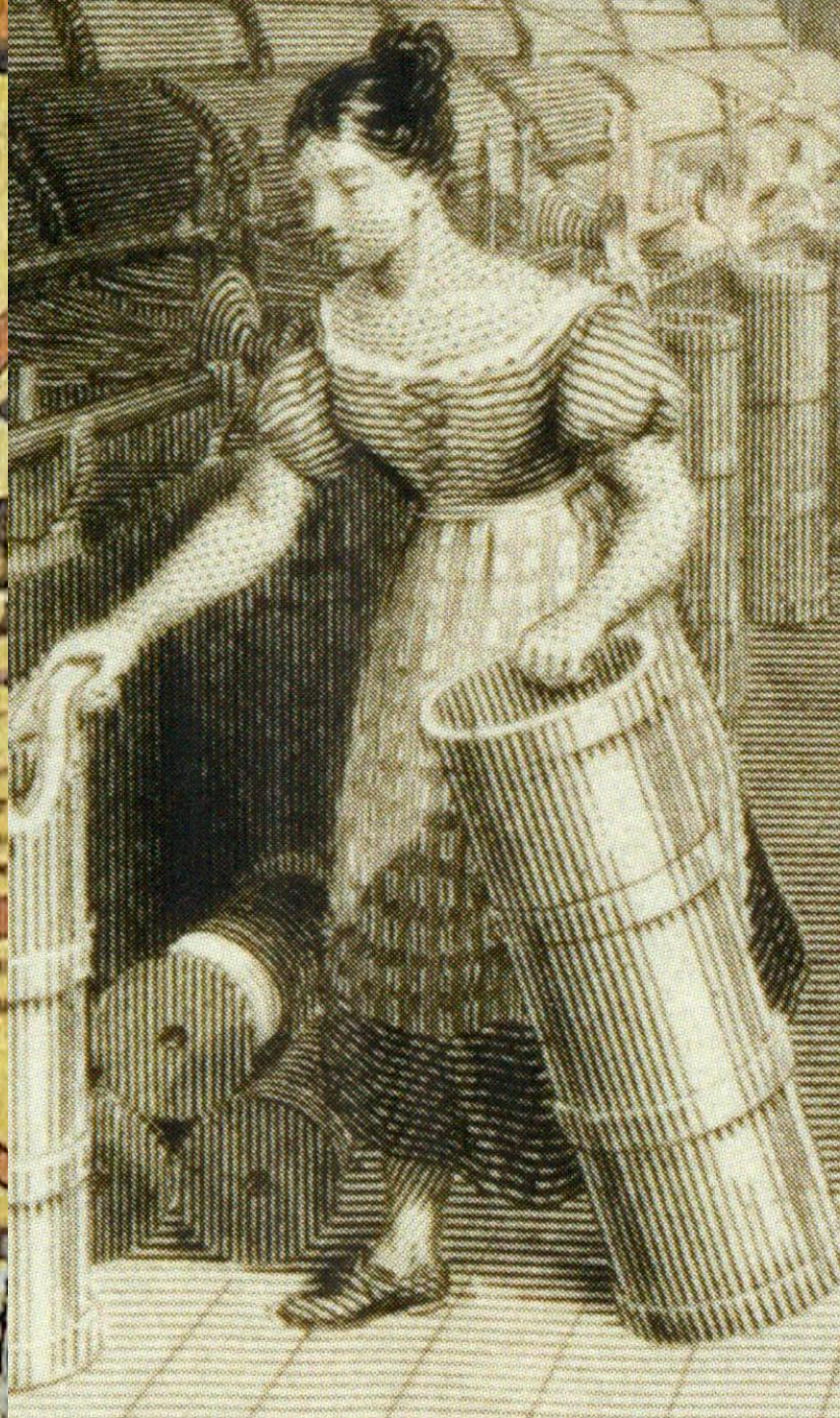
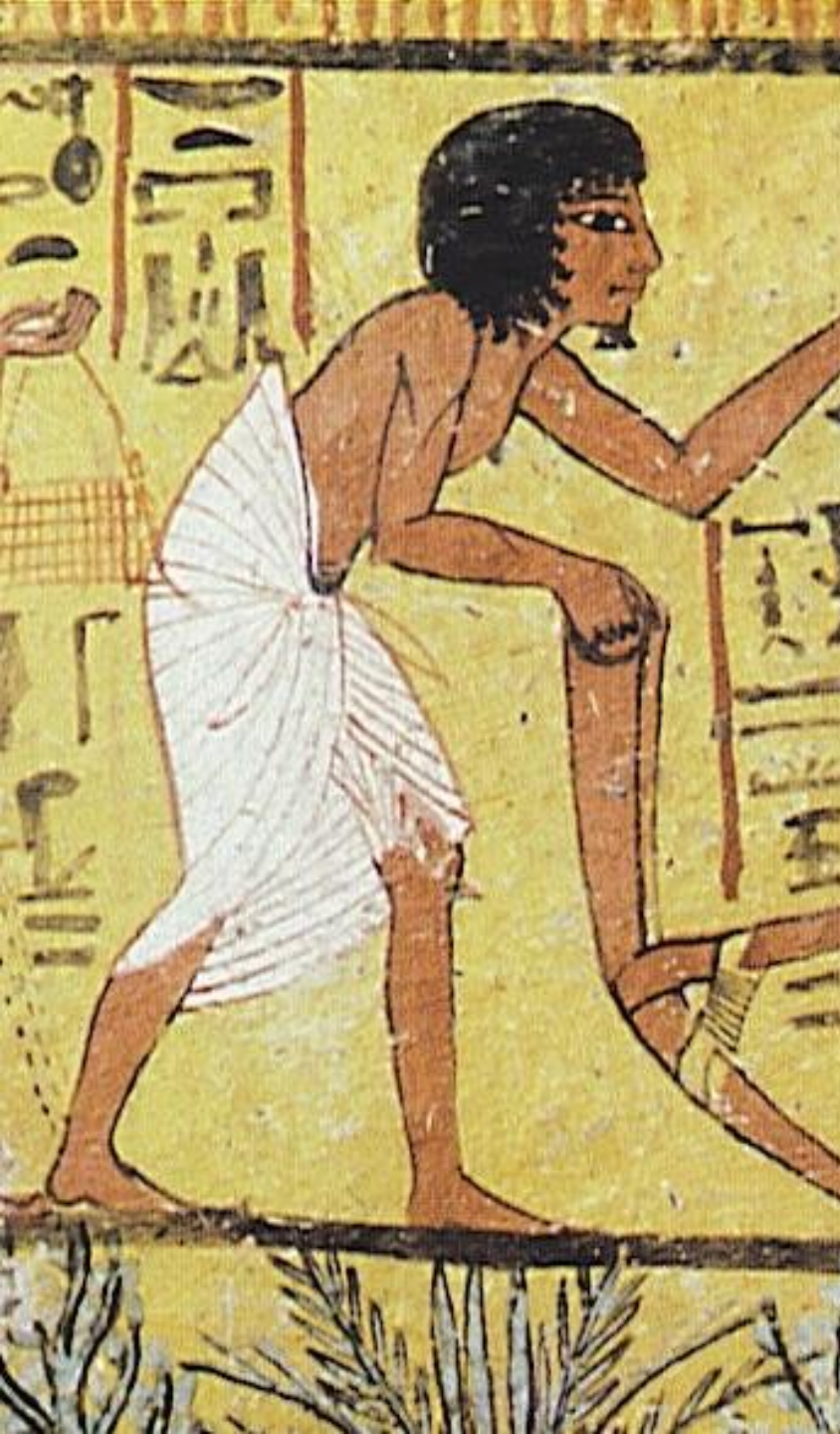
With Great Power Comes Great Responsibility

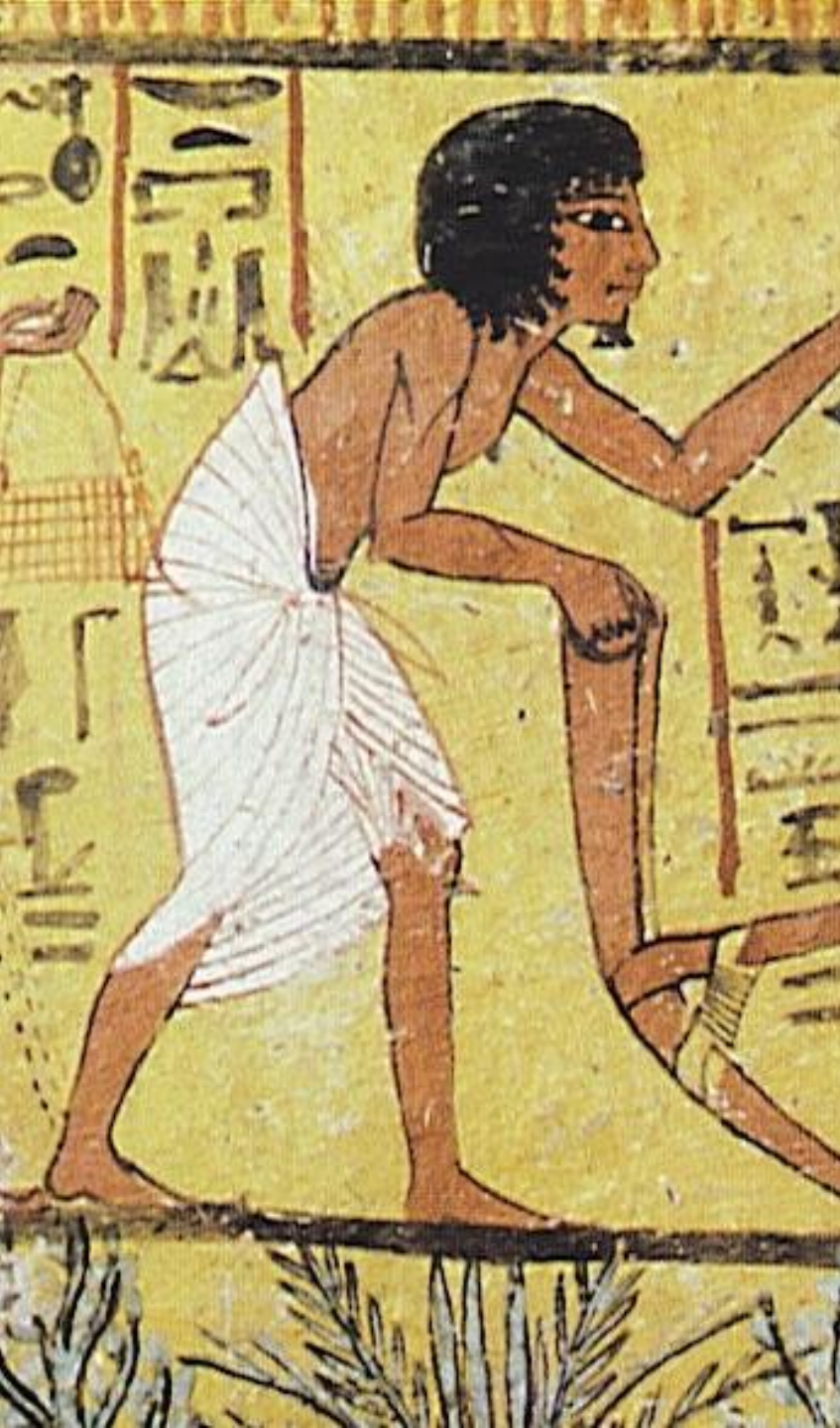
Matthew Renze

Renze Consulting

“With great power comes
great responsibility.”

-- Uncle Ben





Agricultural Superpower

Large armies

Abuses

Acquire land
Enslave labor
Subjugation

Industrial Superpower

Mechanized warfare



Abuses

Control resources
Maintain colonies
Expand influence

**Information
Superpower**

Cyber warfare

Abuses

Mass surveillance
Industrial sabotage
Political propaganda





Purpose

To learn how to use AI
responsibly and ethically

Overview

1. AI Ethics for Today



Overview

1. AI Ethics for Today
2. AI Ethics for Tomorrow



Overview

1. AI Ethics for Today
2. AI Ethics for Tomorrow
3. AI Ethics for The Future

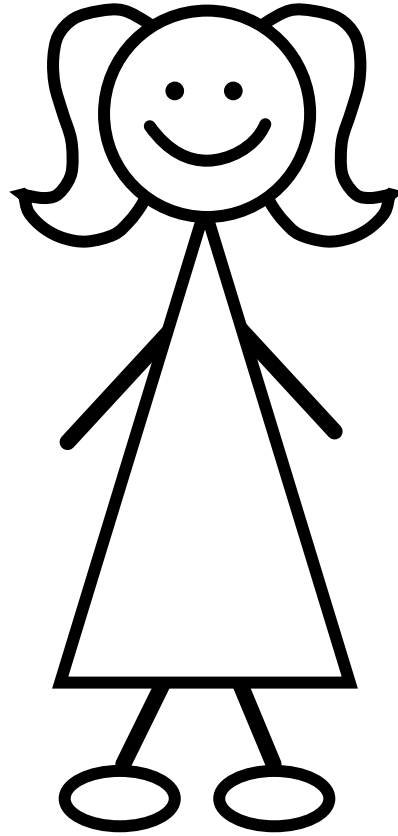
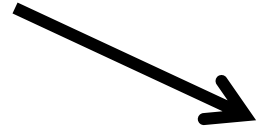


AI Ethics for Today

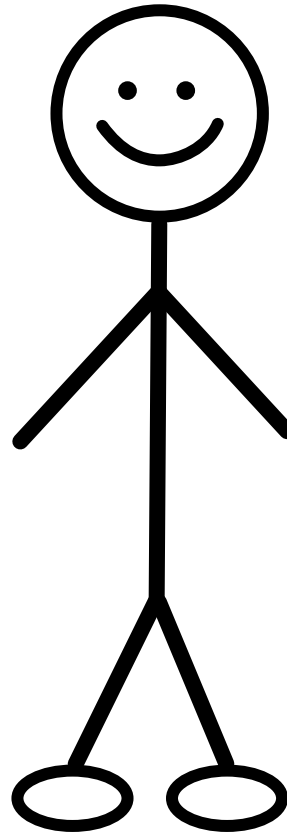
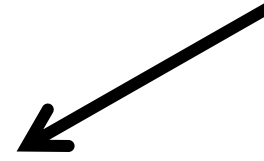
Alice and Bob in AI Land

Chapter 1:
"AI Ethics of Today"

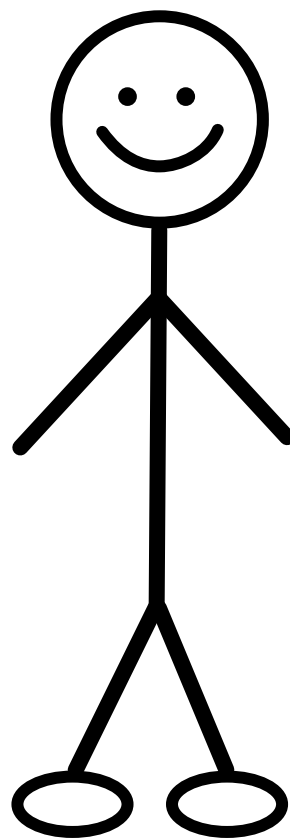
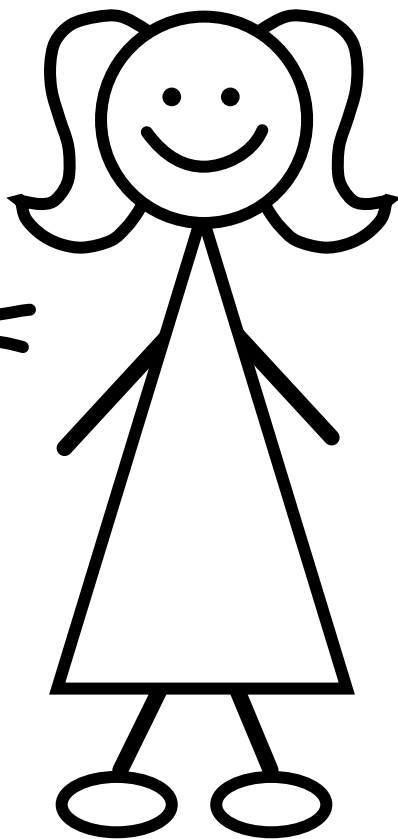
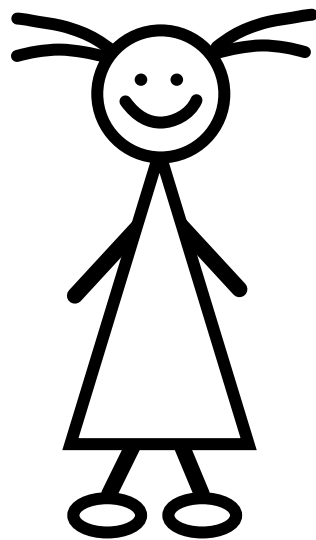
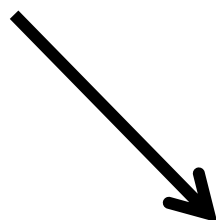
Alice



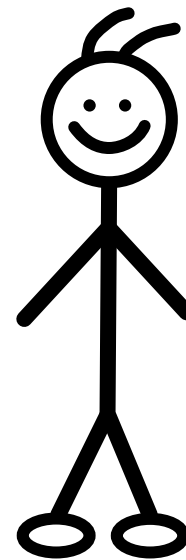
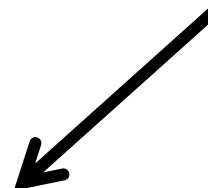
Bob

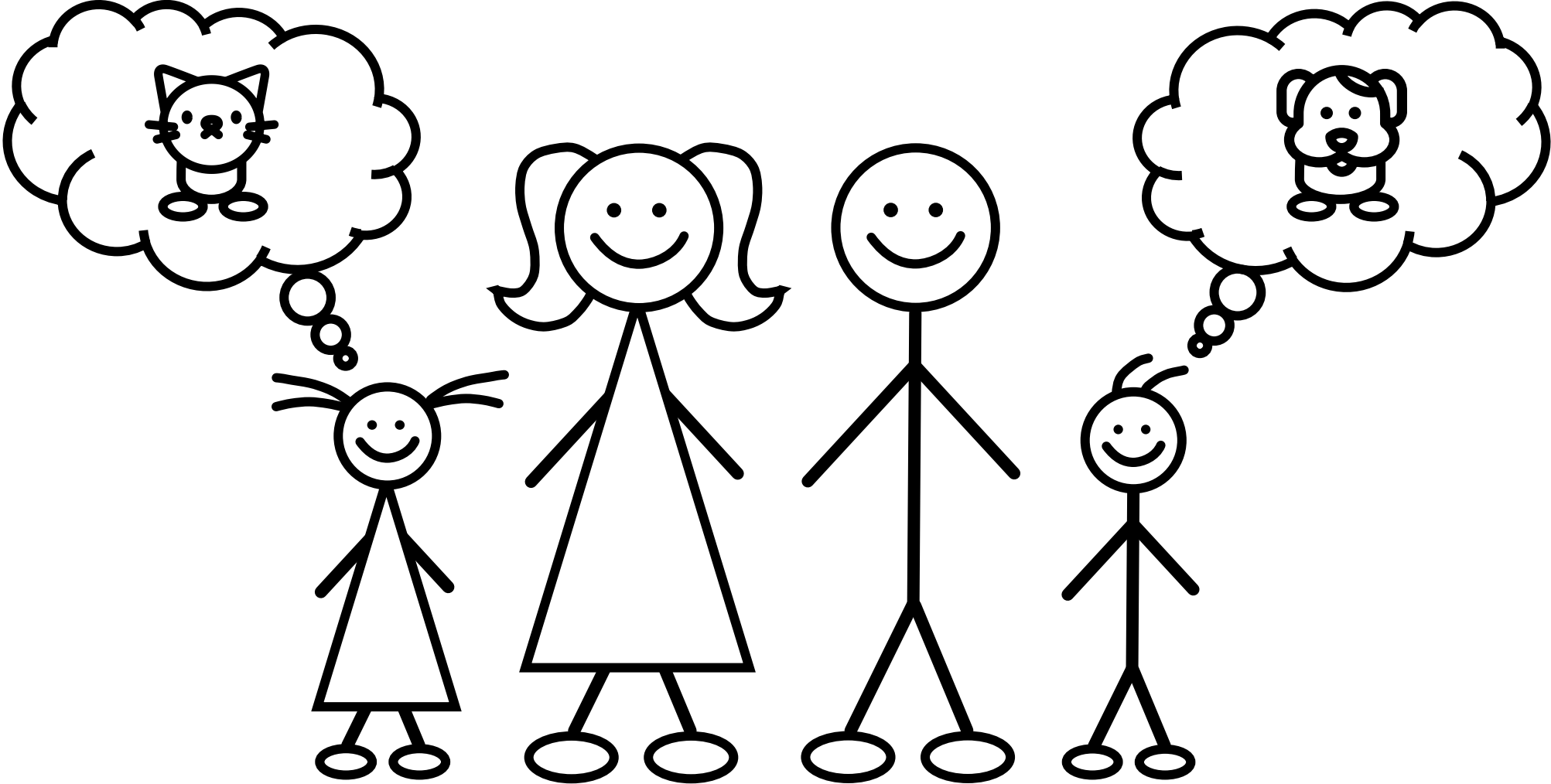


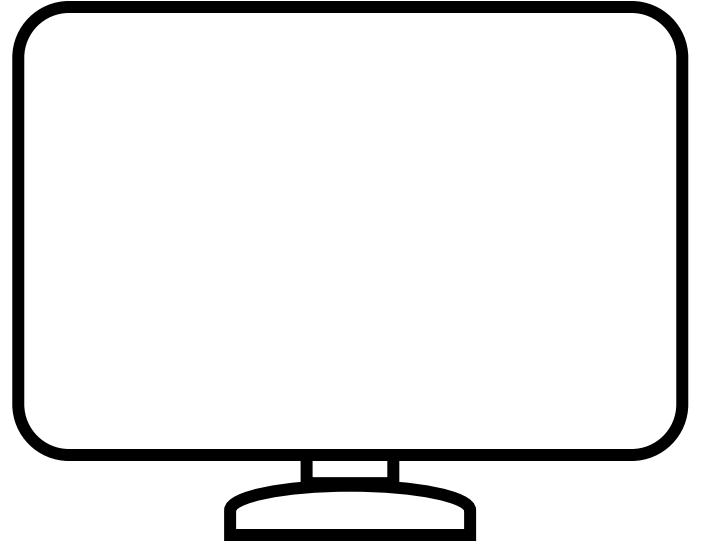
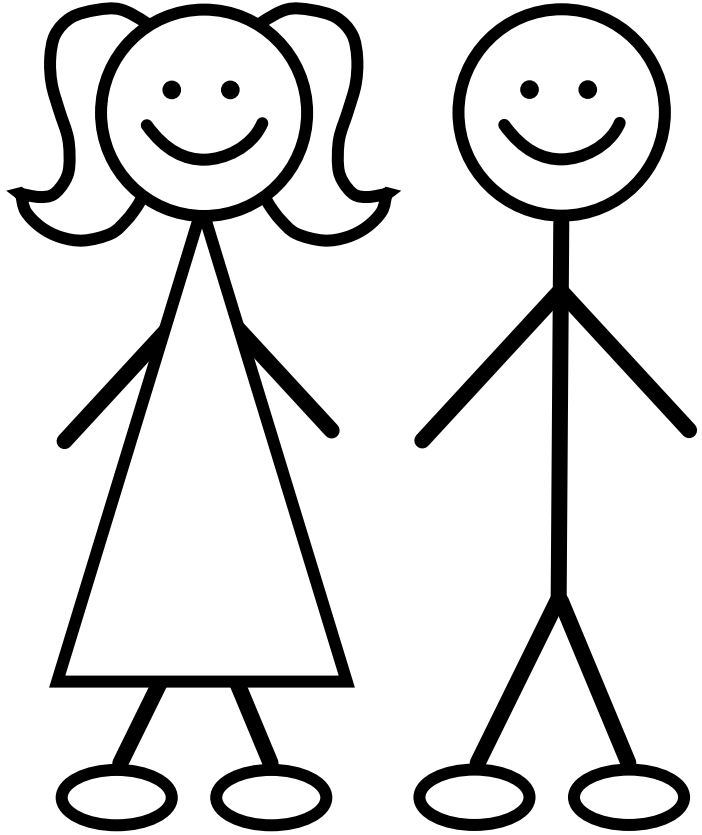
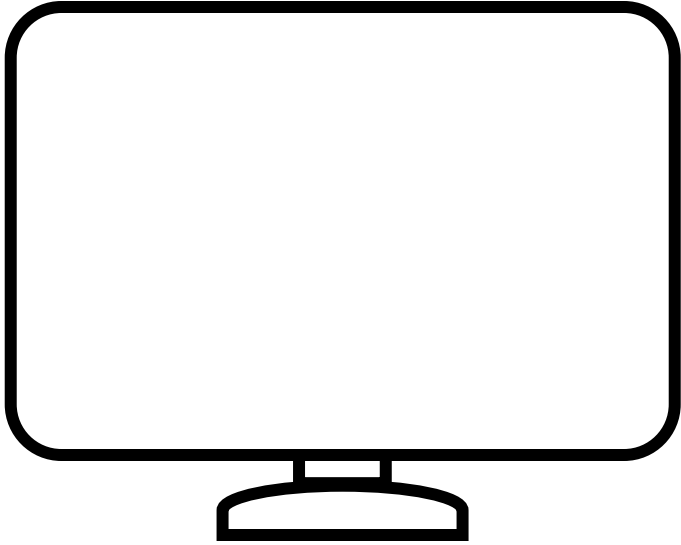
Claire

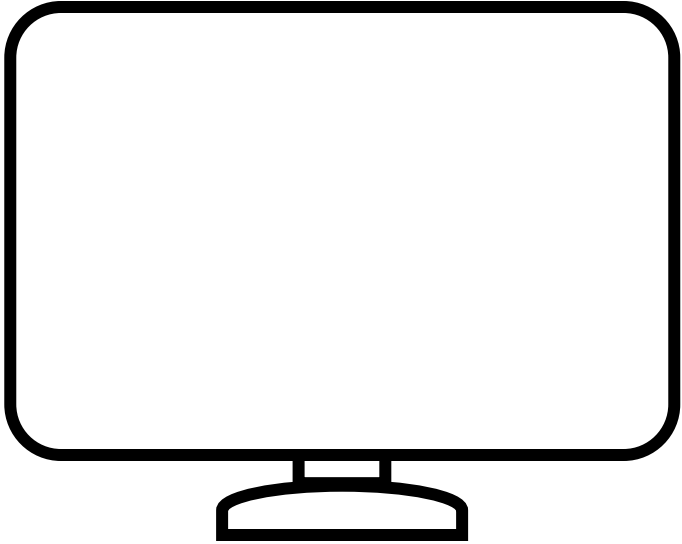
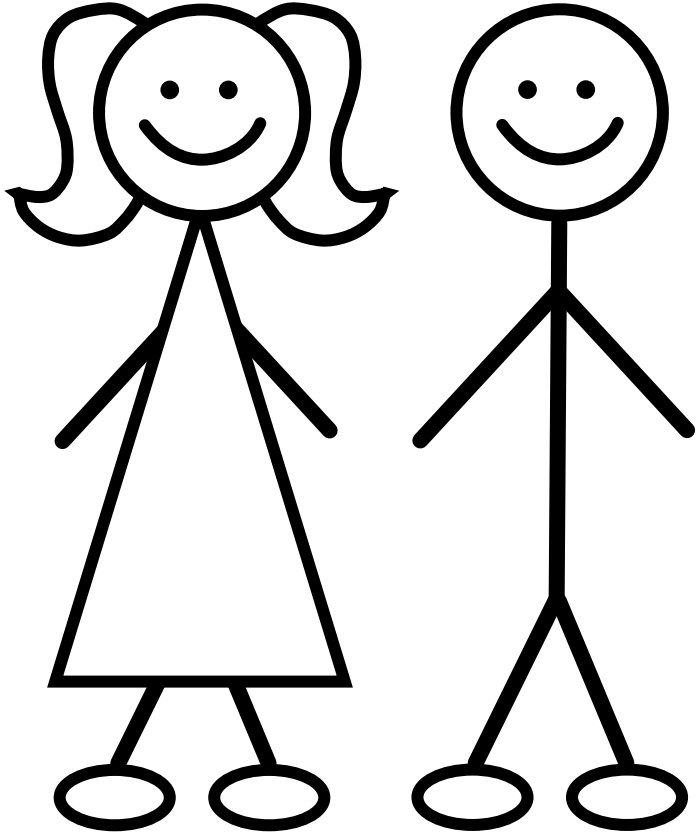
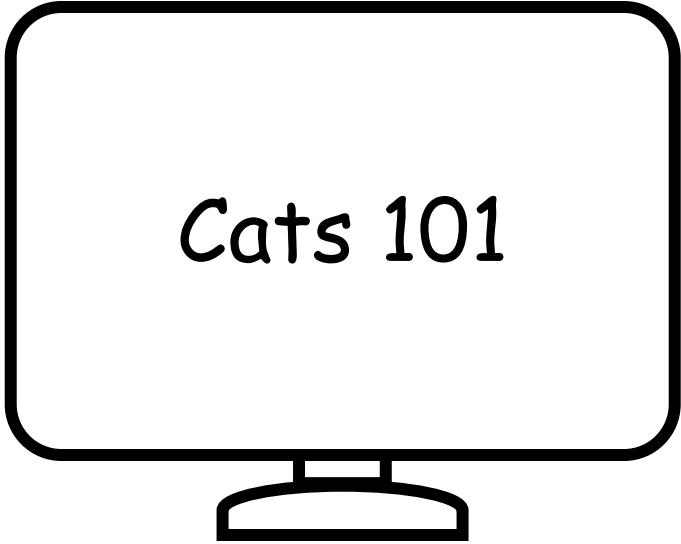


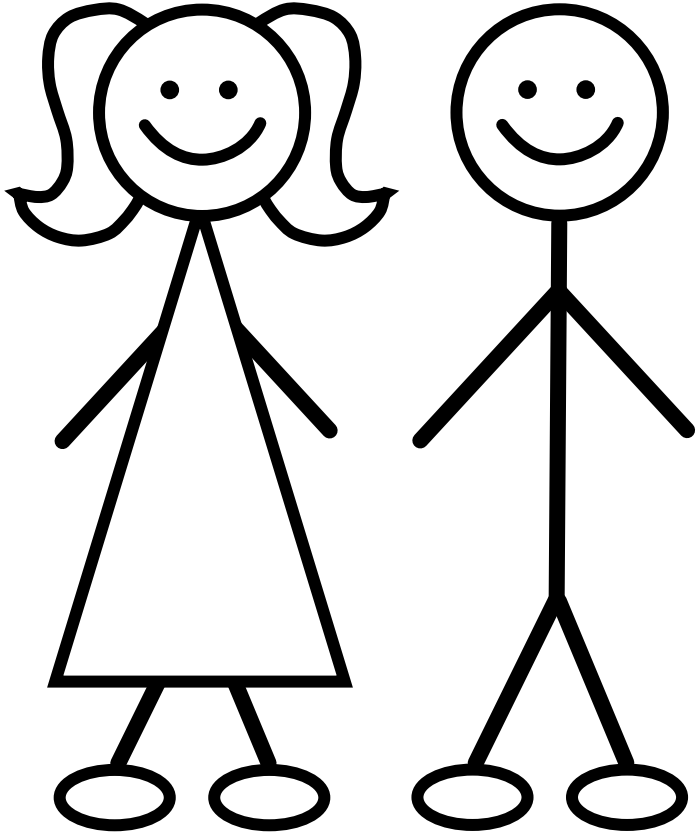
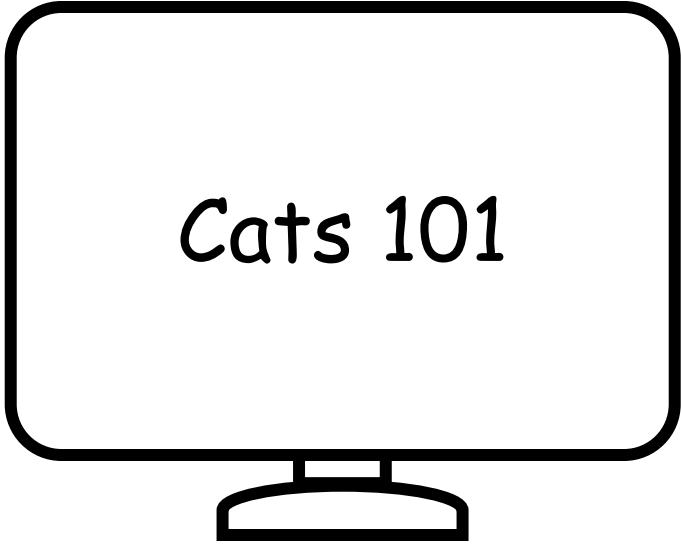
Dan

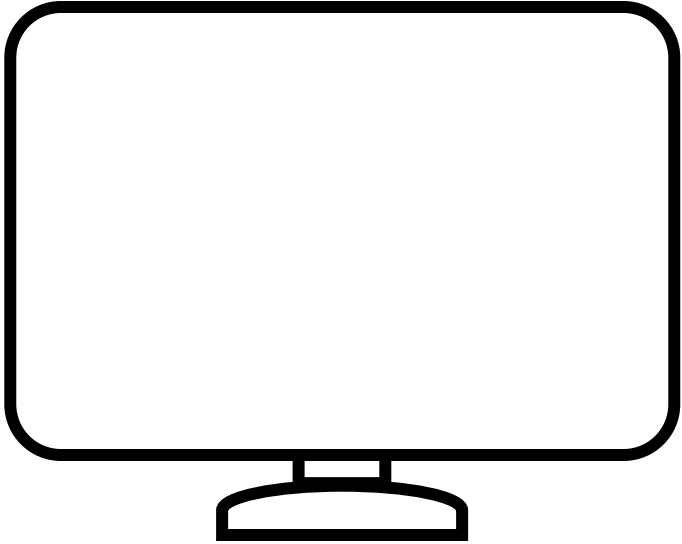
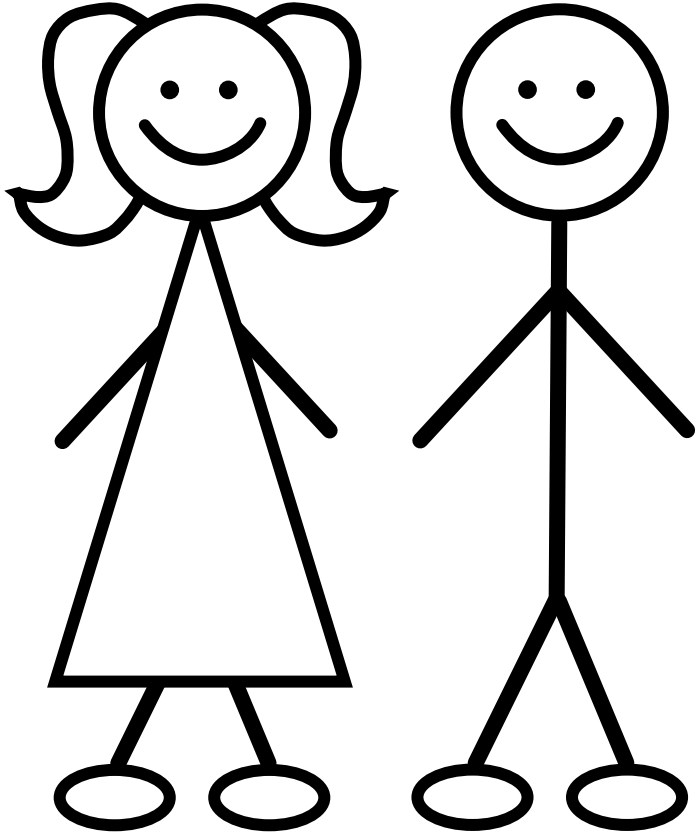


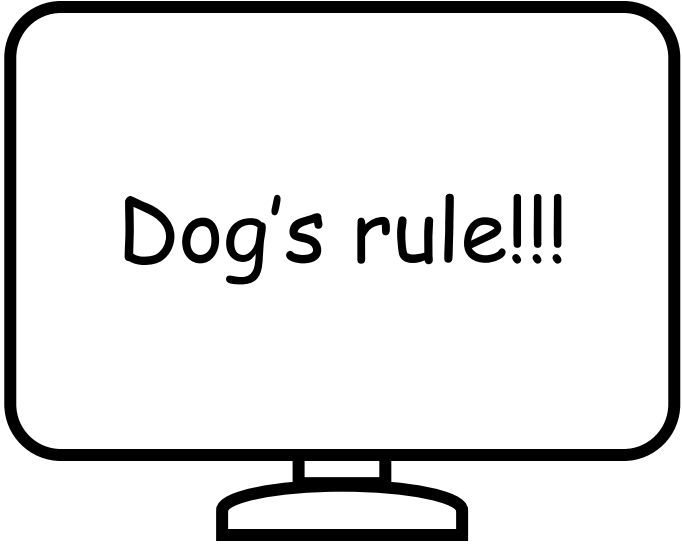
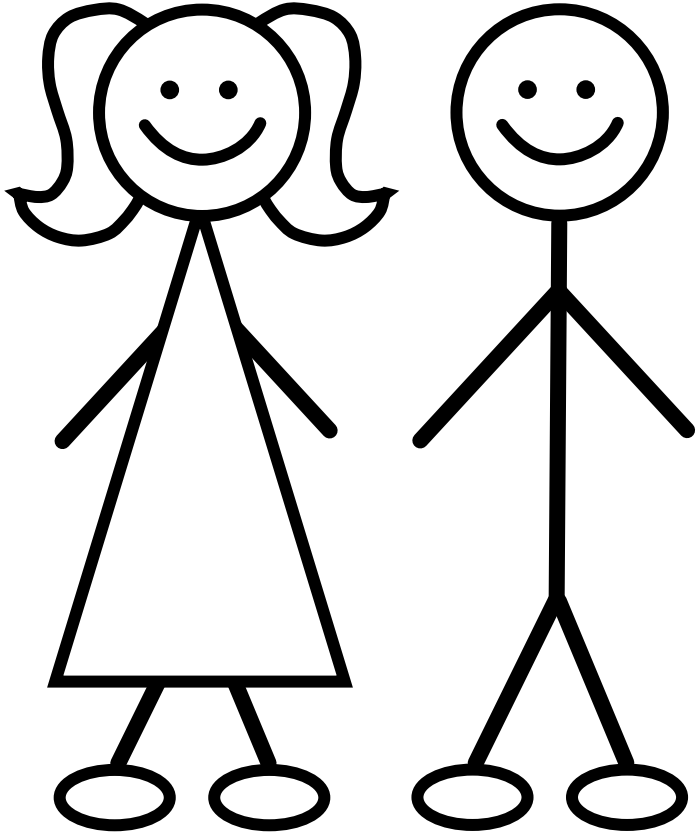


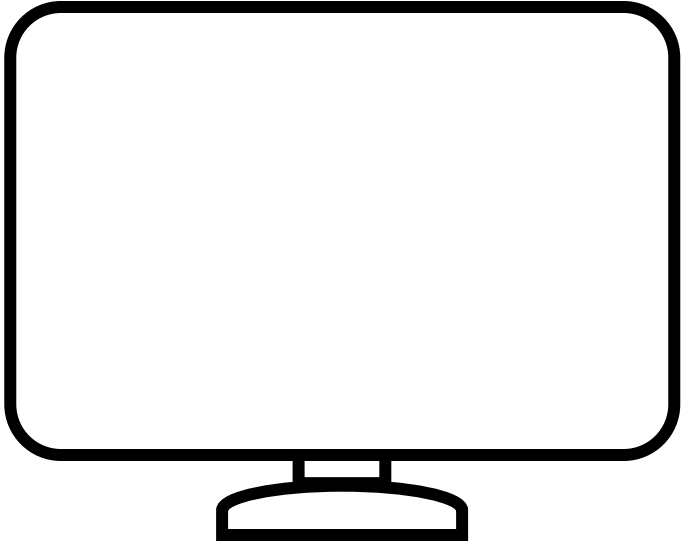
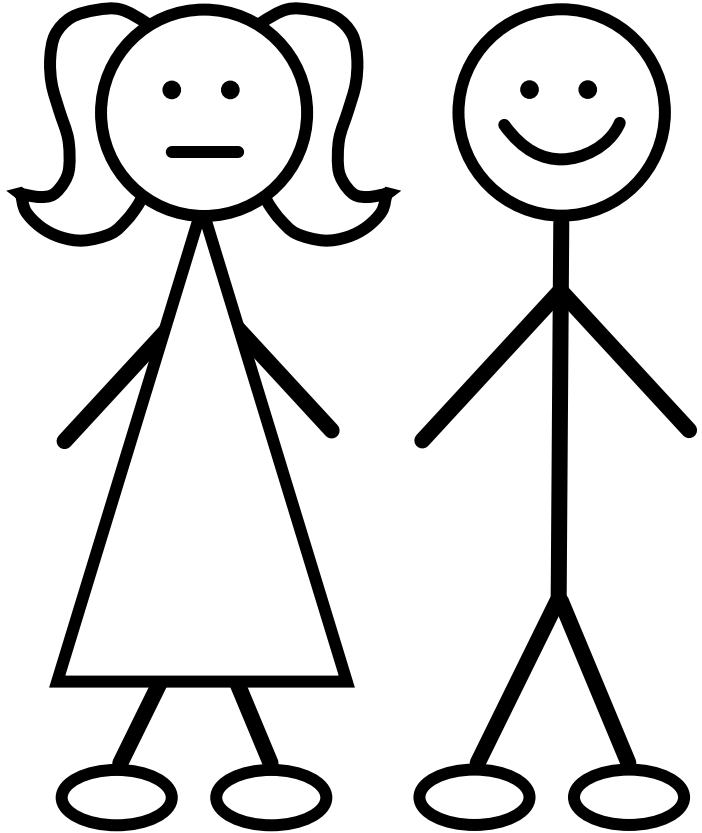




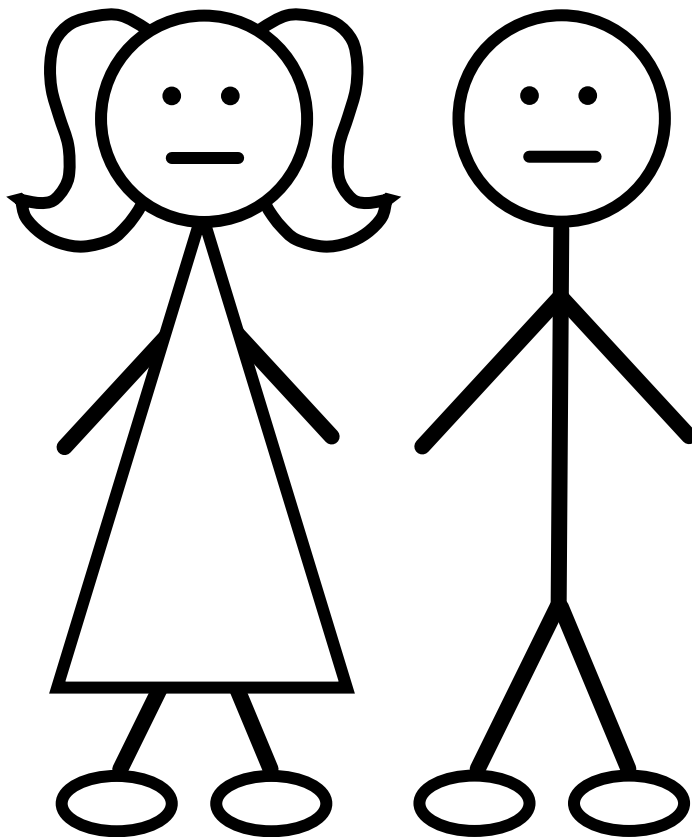




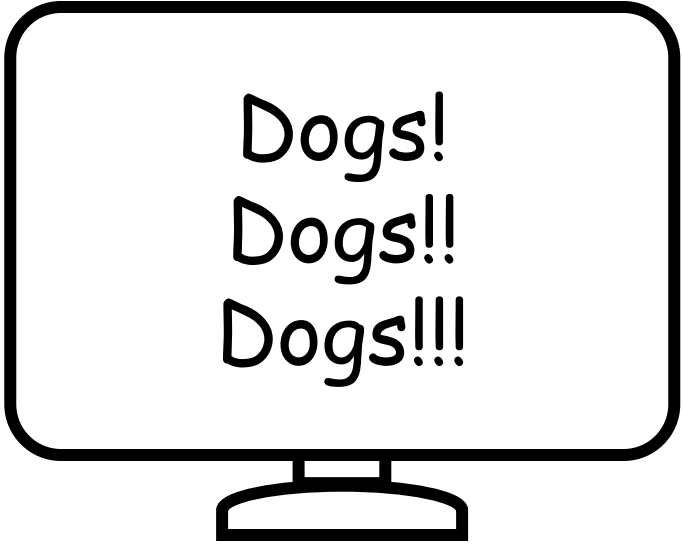
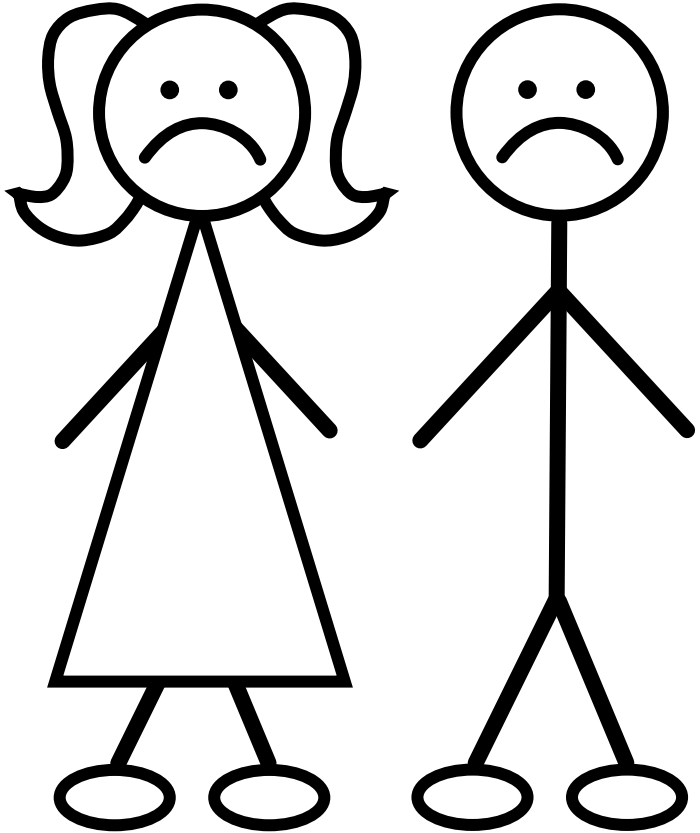
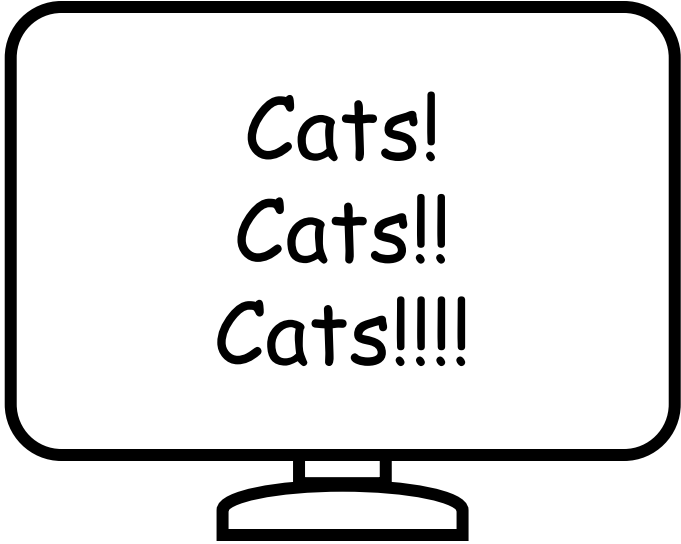


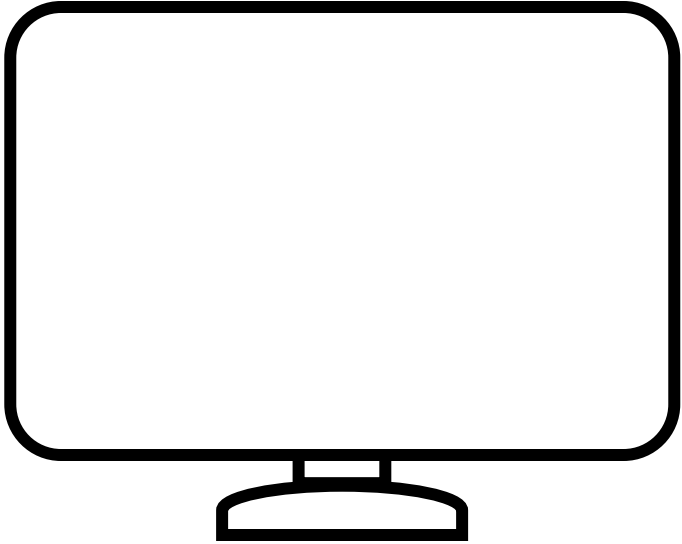
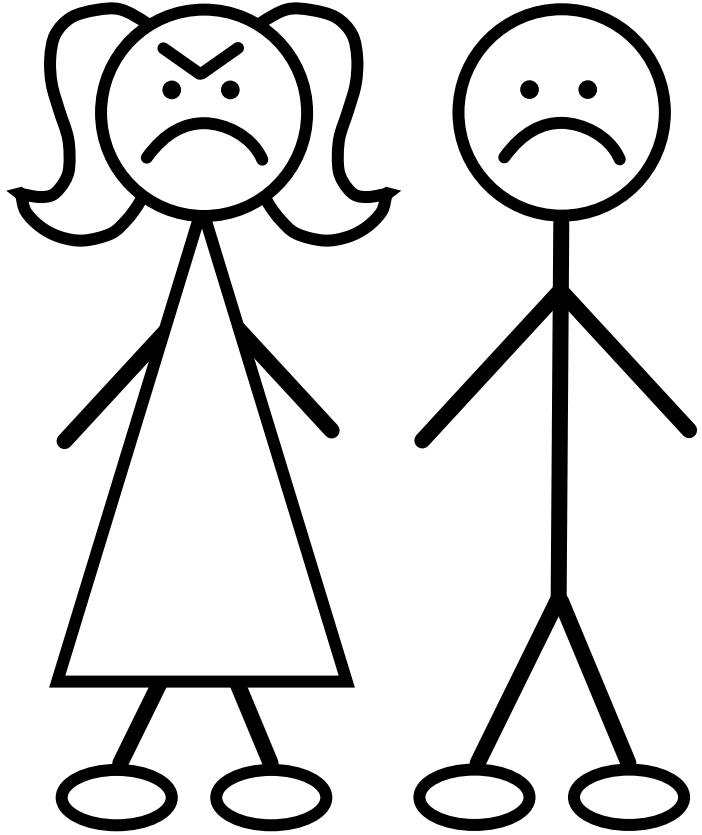


10 reasons I
hate dogs! :(

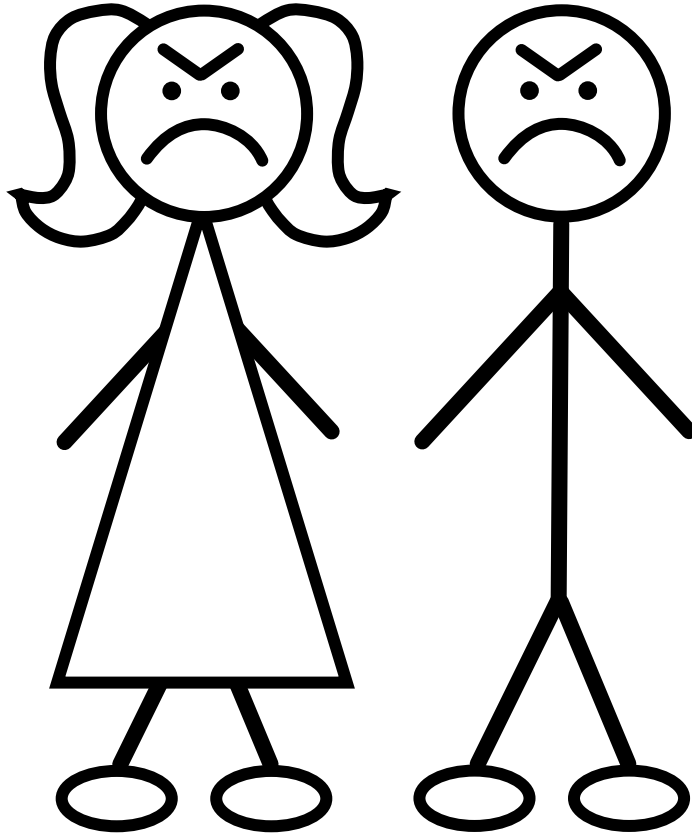


Cats are total
weirdos! :(

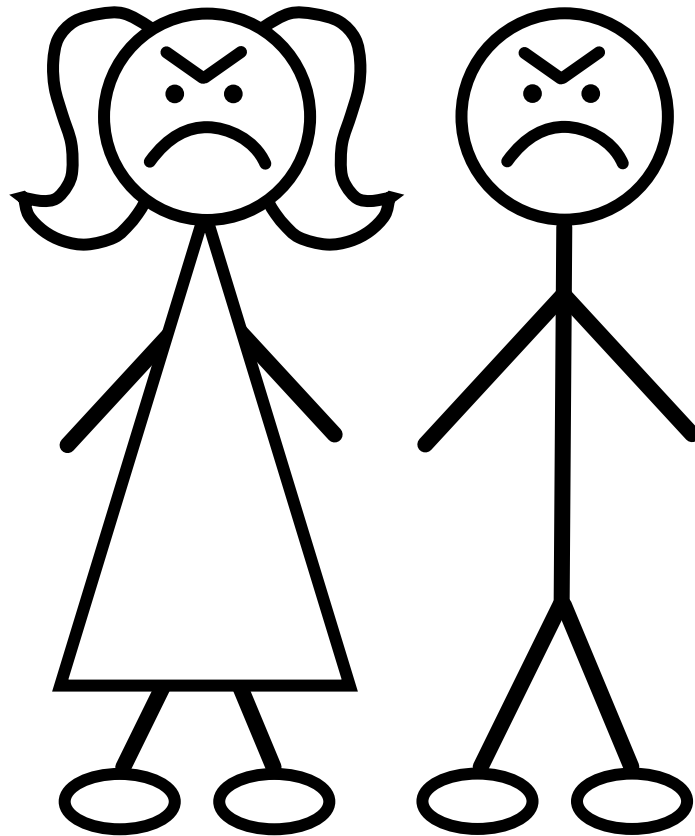


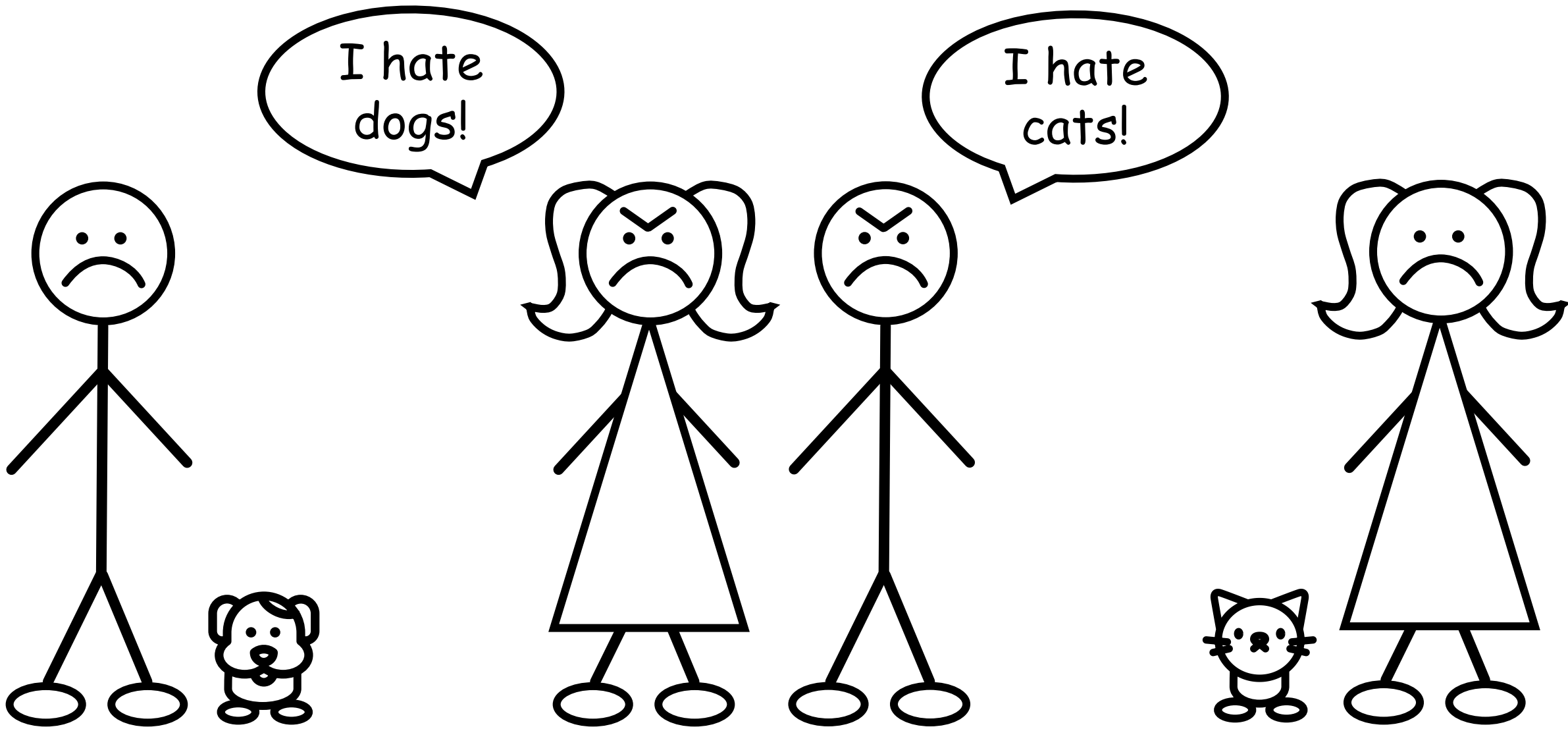


Dogs are
horrible!!! >:(



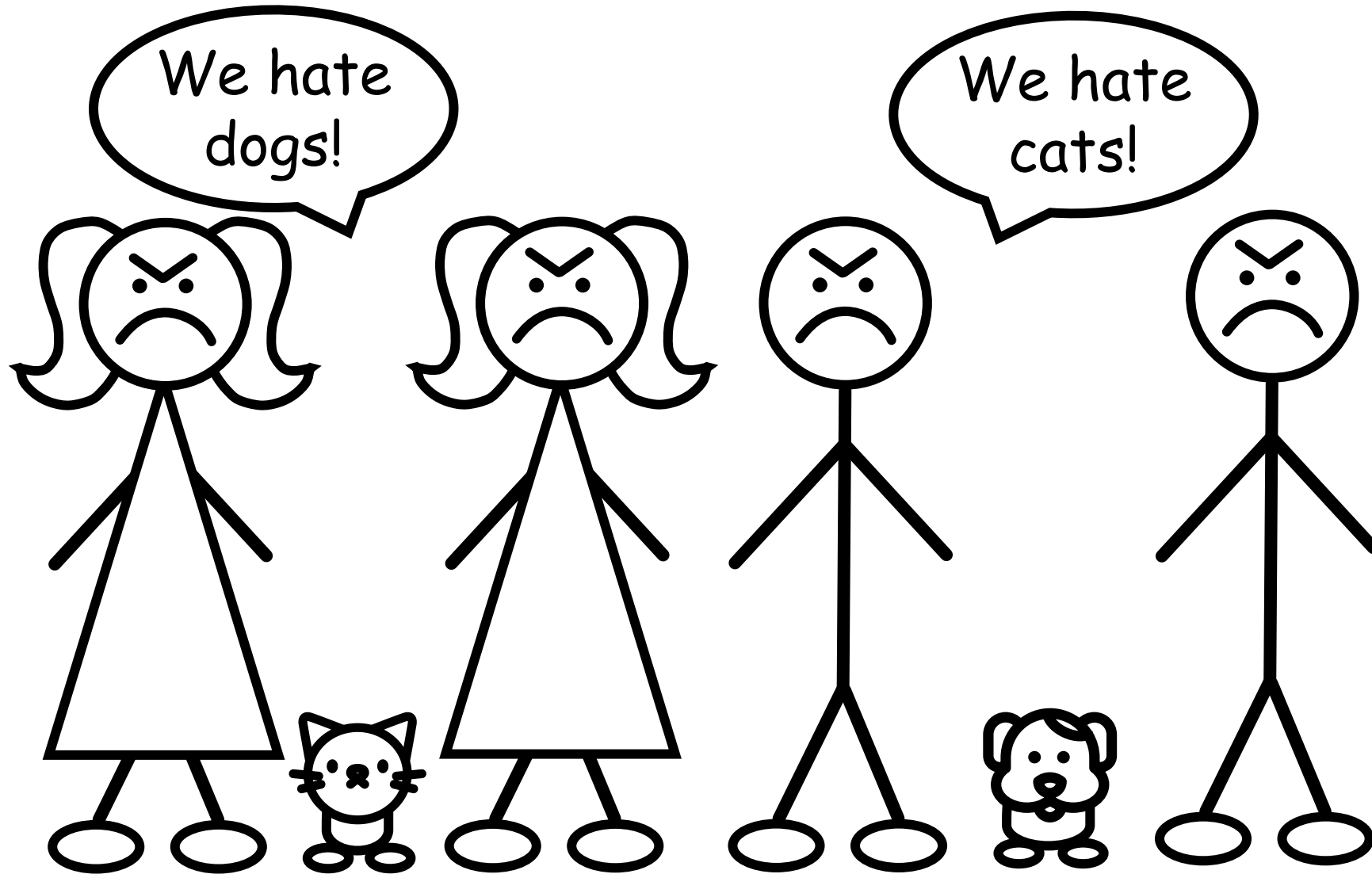
Cats are
terrible!!! >:(





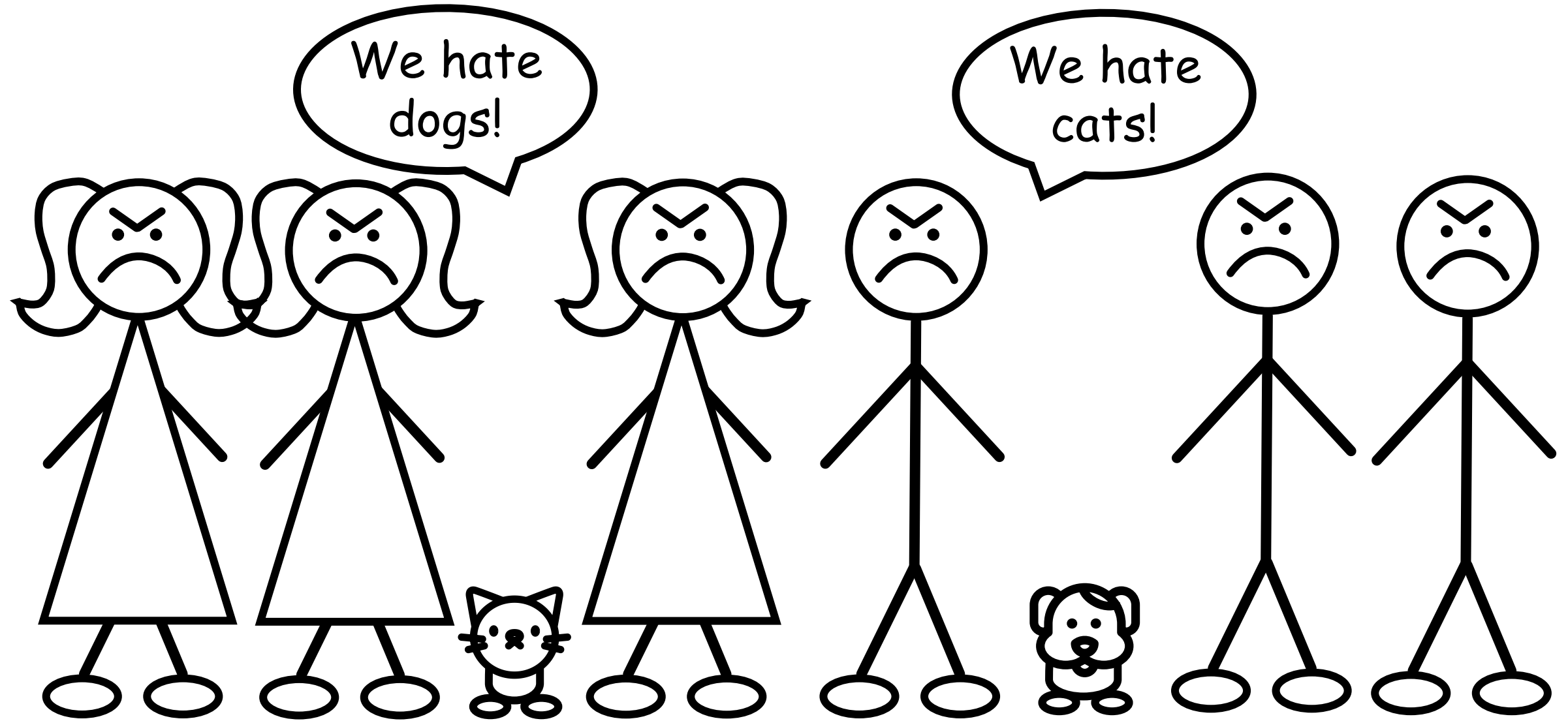
I hate
dogs!

I hate
cats!



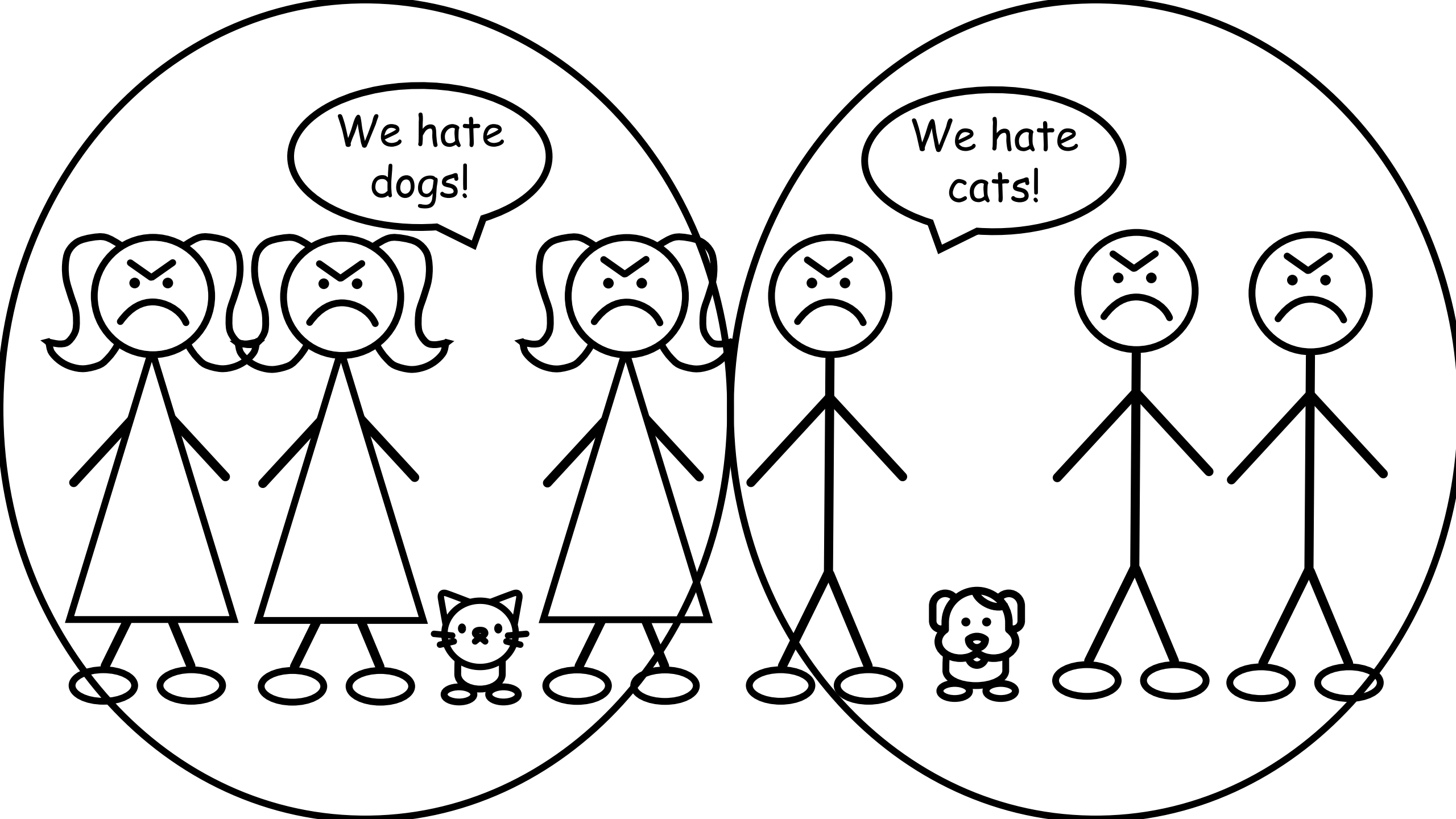
We hate
dogs!

We hate
cats!



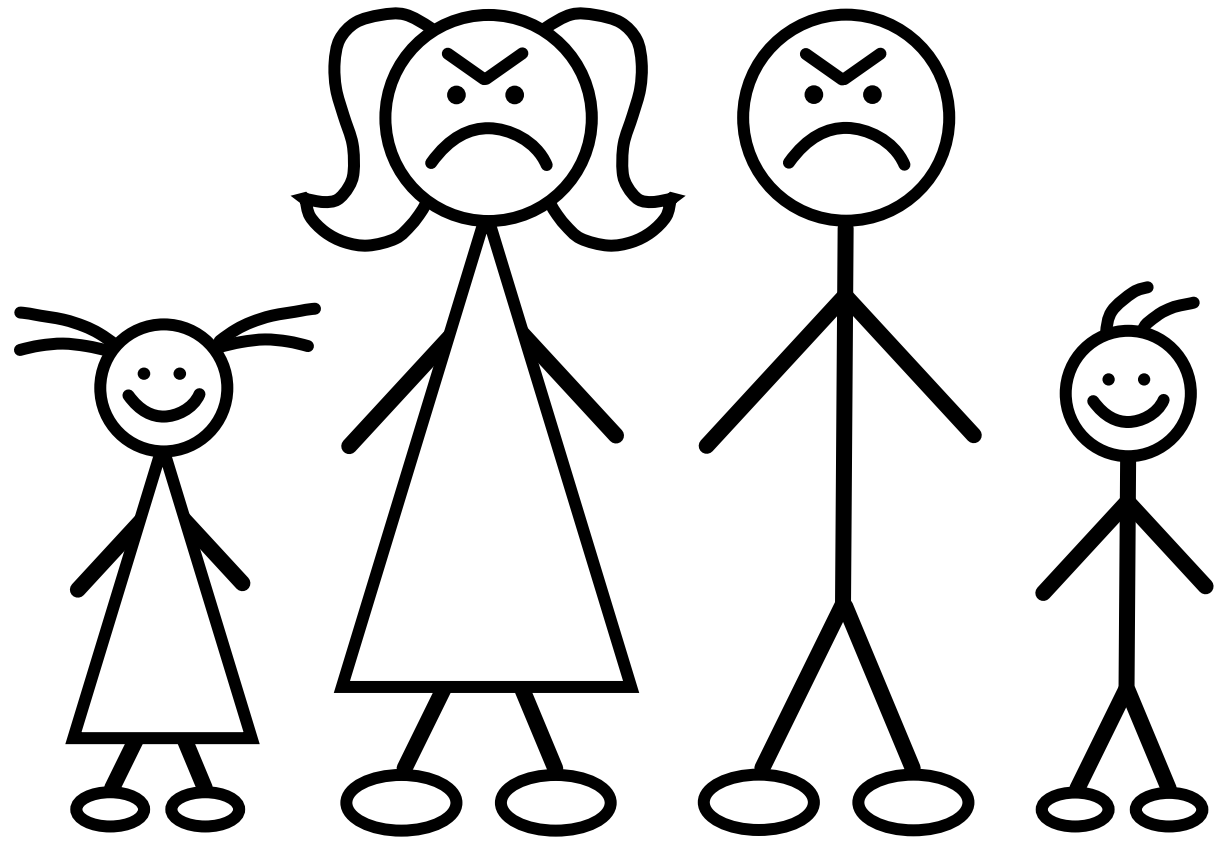
We hate
dogs!

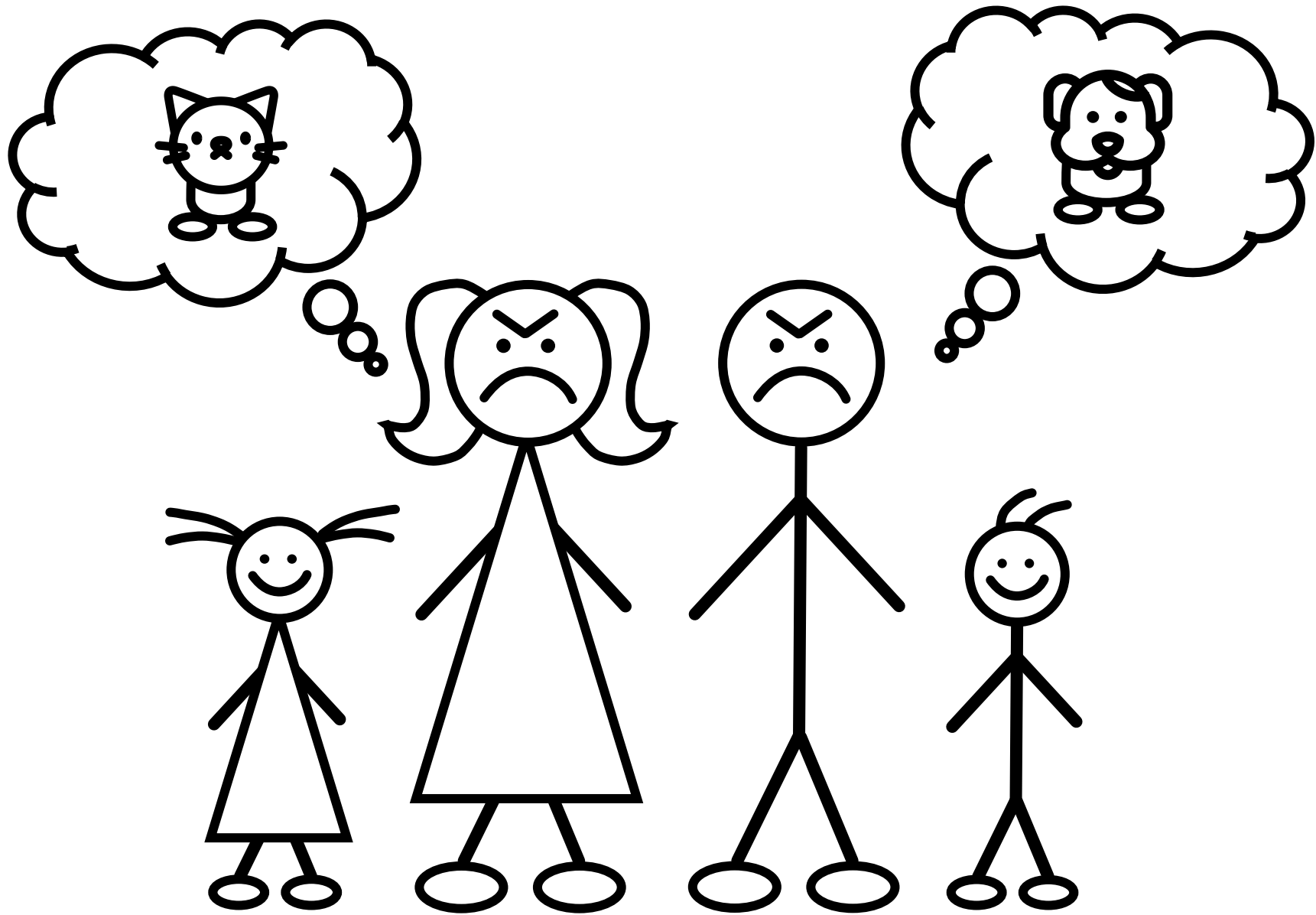
We hate
cats!



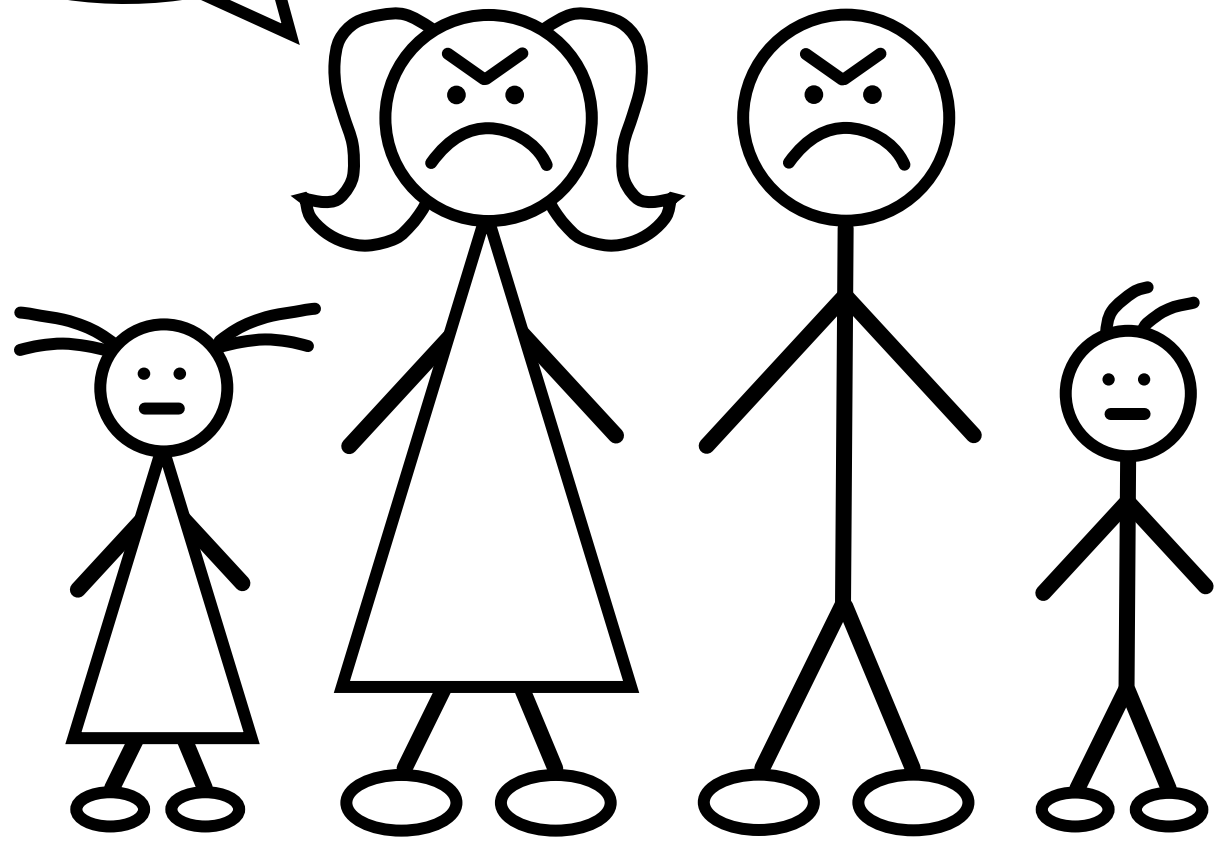
We hate dogs!

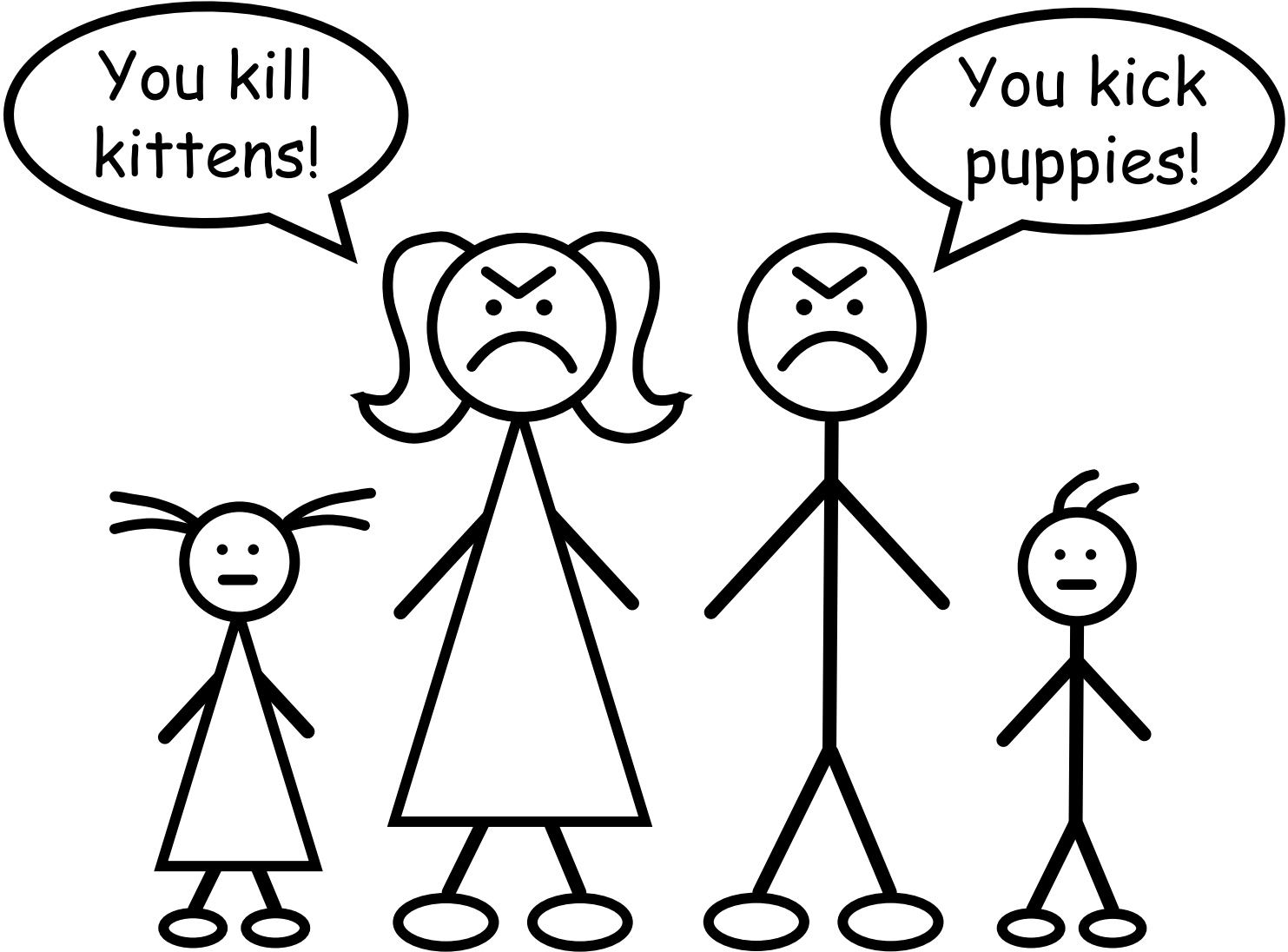
We hate cats!

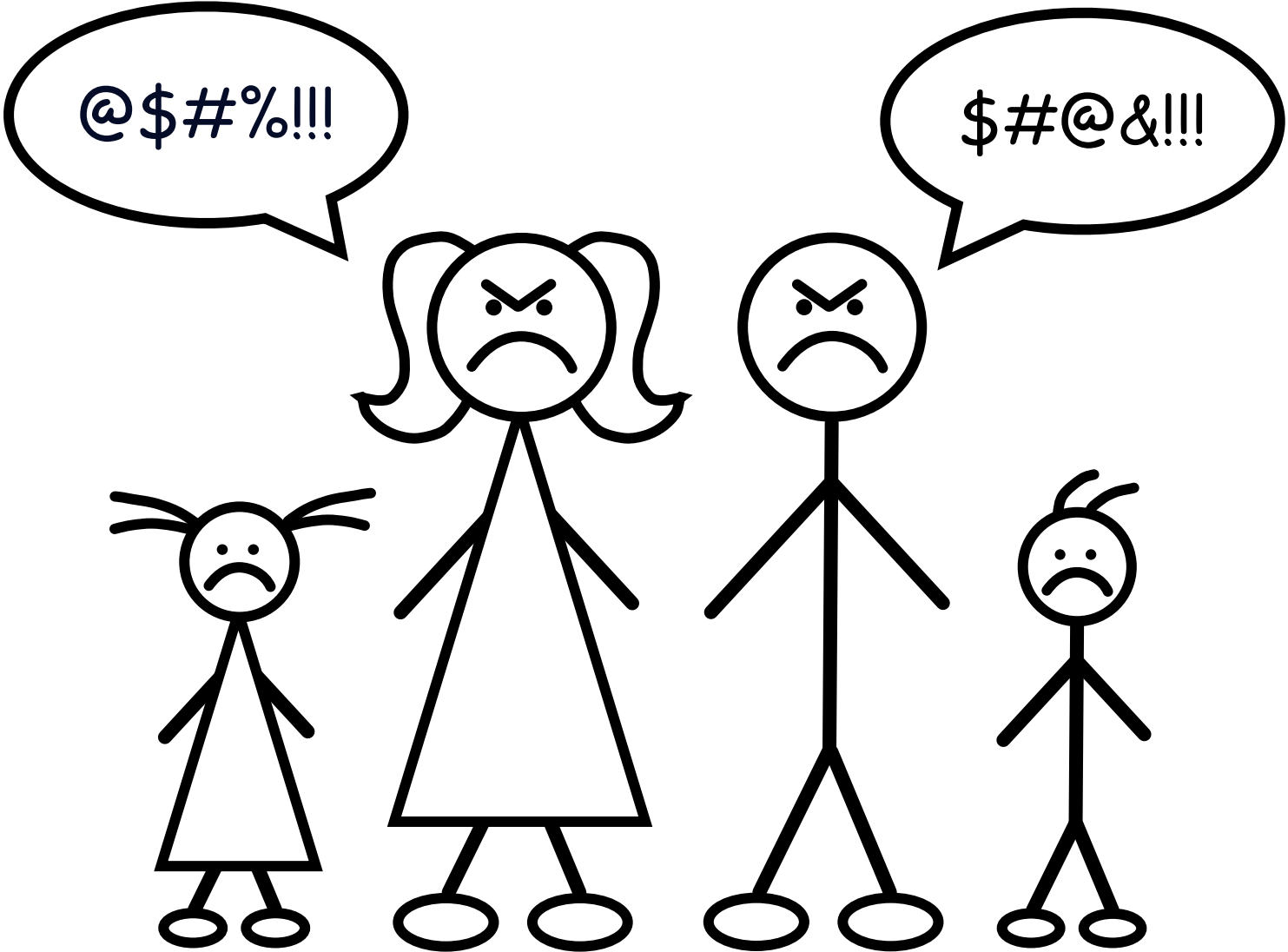




You kill kittens!

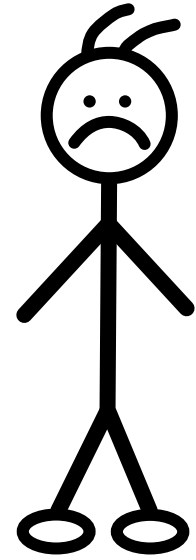
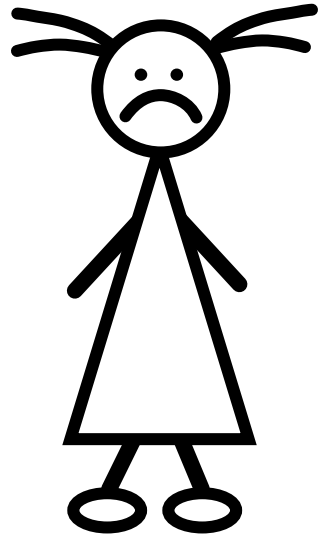






@\$#%!!!

\$#@&!!!



Benefits

TV Shows

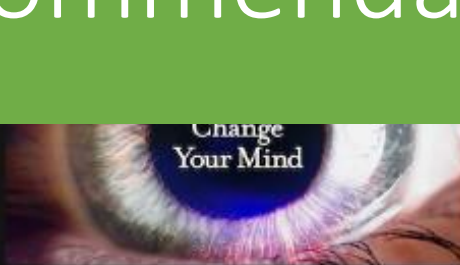
Genres



Suggestions For You



Recommendation



A woman and a young girl are sitting on a light-colored sofa in a bright, airy room. They are both looking towards a dark grey smart speaker with a glowing blue ring at the top, which is positioned in the foreground. The woman is smiling warmly, and the girl is looking intently at the device. The background shows a window with sheer curtains and a potted plant on the right side.

Voice assistants



Robot vacuums



Navigation software

A person is seen from behind, wearing a white t-shirt and large black headphones. They are sitting in front of a large monitor displaying a colorful, action-packed video game. The game's interface includes various HUD elements like health bars, a mini-map, and character names. The scene is dimly lit, with the primary light source being the monitor's glow.

Video games



Model: GPT-4



What are the top 10 benefits of modern AI?



ChatGPT

10 benefits of modern AI:

1. Improved decision-making: AI systems can analyze vast amounts of data quickly and accurately, allowing for better and faster decision-making in fields like finance, healthcare, and manufacturing.



Other Benefits

Automation

Enhanced
productivity

Improved
decision-making

Personalization

Cost savings

Risk mitigation

Predictive
maintenance

Enhanced
cybersecurity

Accessibility

Problems

A person wearing a grey hoodie is sitting at a desk, viewed from behind. They are looking at a laptop screen. The scene is dimly lit, with the person's face obscured by the hood. An orange horizontal bar is overlaid across the middle of the image, containing the text "Recommendation rabbit holes" in white.

Recommendation rabbit holes



Maximize
engagement

Recommendation rabbit holes

Maximize
engagement

More extreme
content

Recommendation rabbit holes

The background of the slide is a dark, low-key photograph of a person sitting at a desk, viewed from behind. The person is wearing a dark jacket and is looking at a laptop screen. The lighting is dramatic, highlighting the person's shoulders and the back of their head against a black background.

Maximize
engagement

More extreme
content

Information
bubbles

Recommendation rabbit holes

Maximize
engagement

More extreme
content

Information
bubbles

Recommendation rabbit holes

Echo
chambers

Maximize
engagement

More extreme
content

Information
bubbles

Recommendation rabbit holes

Echo
chambers

Conspiratorial
thinking

Maximize
engagement

More extreme
content

Information
bubbles

Recommendation rabbit holes

Echo
chambers

Conspiratorial
thinking

Internet
addictions



ID 432463743777



ID 984093591409



ID 435236626545



ID 432625666326

Lack of privacy

[MATCH]



John Doe

AGE 35 HEIGHT 5'11"

Occupation **Manager**

Interests **Technology, VR, Travel**

Location **London**

AI-enabled surveillance systems



ID 432463743777



ID 984093591409



ID 435236626545



ID 432625666326

Lack of privacy

[MATCH]



John Doe

AGE 35 HEIGHT 5'11"

Occupation Manager

Interests Technology, VR, Travel

Location London

AI-enabled surveillance systems

Social media predicts private information



093591409



ID 435236626545



ID 432625666326

Lack of privacy

[MATCH]



John Doe

AGE 35 HEIGHT 5'11"

Occupation Manager

Interests Technology, VR, Travel

Location London

AI-enabled surveillance systems

Social media predicts private information

Shopping data predicts pregnancy

Lack of privacy

[MATCH]



John Doe

AGE 35

HEIGHT 5'11"

Occupation Manager

Interests Technology, VR, Travel

Location London

AI-enabled surveillance systems

Social media predicts private information

Shopping data predicts pregnancy

Lack of privacy

Deanonymization (DOB, sex, zip code)

[MATCH]



John Doe

AGE 35 HEIGHT 5'11"

Occupation Manager

Interests Technology, VR, Travel

Location London

AI-enabled surveillance systems

Social media predicts private information

Shopping data predicts pregnancy

Lack of privacy

Deanonymization (DOB, sex, zip code)

What is privacy with AI surveillance?

John Doe	
AGE 35	HEIGHT 5'11"
Occupation	Manager
Interests	Technology, VR, Travel
Location	London

AI-enabled surveillance systems

Social media predicts private information

Shopping data predicts pregnancy

Lack of privacy

Deanonymization (DOB, sex, zip code)

What is privacy with AI surveillance?

Little privacy now, soon much less

Algorithmic bias



Racially biased
criminal risk scores

Algorithmic bias



Racially biased
criminal risk scores

Gender biased
resume screening

Algorithmic bias



Racially biased
criminal risk scores

Gender biased
resume screening

Dark skin biased
facial recognition

Algorithmic bias



Racially biased
criminal risk scores

Gender biased
resume screening

Dark skin biased
facial recognition

Algorithmic bias

Beauty filters
lighten dark skin



Racially biased
criminal risk scores

Gender biased
resume screening

Dark skin biased
facial recognition

Algorithmic bias

Beauty filters
lighten dark skin

Human bias in
training data



Racially biased
criminal risk scores

Gender biased
resume screening

Dark skin biased
facial recognition

Algorithmic bias

Beauty filters
lighten dark skin

Human bias in
training data

Reinforces existing
social stereotypes



Lack of trust

Lack of
transparency

Lack of trust

Lack of
transparency

Lack of
interpretability

Lack of trust

Lack of
transparency

Lack of
interpretability

Lack of trust

Adversarial
attacks

Lack of
transparency

Lack of
interpretability

Lack of trust

Adversarial
attacks

Uncanny
valley



Hallucination

**FAKE
NEWS**

Misinformation



et non sunt ipsum dolor in hendrerit in vulputate velit esse
consectetur adipisicing elit, sed do eiusmod tempor incididunt ut
labore et dolore magna aliqua. Ut enim ad minim veniam, quis
nostrud exercitation ullamco laboris nisi ut aliquip ex ea
commodo consequat. Duis aute irure dolor in reprehenderit in
vulputate velit esse mollit anim id est laborum.

A man with dark hair, a beard, and glasses is sitting at a desk in a library. He is looking at a laptop with a confused expression, his hands raised in a gesture of uncertainty. The background is filled with bookshelves. An orange banner is overlaid on the image with the text "Lack of AI literacy".

Lack of AI literacy

Solutions

A close-up, angled view of a smartphone screen. The top portion shows a search bar with the text "Search or type URL" and a tab indicator showing "1". Below the search bar, a circular icon is visible. The main part of the screen displays text about incognito tabs. A blue horizontal bar is overlaid on the screen, containing the title "Escape information bubbles".

Escape information bubbles

Pages you view in incognito tabs won't
around in your browser's history
or search history after you close
your incognito tabs. Also, you won't
bookmarks you create in incognito tabs.

Use privacy
tools

Escape information bubbles

Pages you view in incognito tabs won't
around in your browser's history
or search history after you close
your incognito tabs. Also, you can't
bookmarks you create in incognito
How

Use privacy
tools

Block
keywords

Escape information bubbles

Pages you view in incognito tabs won't
around in your browser's history
or search history after you close
your incognito tabs. Also, you won't
bookmarks you create in incognito tabs.

Use privacy
tools

Block
keywords

Stop using
“click holes”

Escape information bubbles

Pages you view in incognito tabs won't
around in your browser's history
or search history after you close
your incognito tabs. Also, you won't
bookmarks you create in incognito tabs.

Use privacy
tools

Block
keywords

Stop using
“click holes”

Escape information bubbles

Provide opt-out
options

Use privacy
tools

Block
keywords

Stop using
“click holes”

Escape information bubbles

Provide opt-out
options

Value users
over profit

Use privacy
tools

Block
keywords

Stop using
“click holes”

Escape information bubbles

Provide opt-out
options

Value users
over profit

Support privacy
regulations



Protect your privacy

The background features a person's hands typing on a laptop keyboard. Overlaid on this are several semi-transparent icons: a globe, a dollar sign, a padlock, a Wi-Fi symbol, a shopping cart, and a magnifying glass. A network of dots and lines is also visible in the upper center.

Use privacy
tools

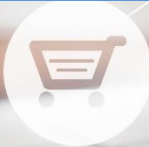
Protect your privacy

Use privacy
tools

Use throw-away
accounts



Protect your privacy

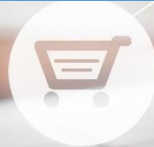


Use privacy
tools

Use throw-away
accounts

Stop using
data harvesters

Protect your privacy



Use privacy
tools

Use throw-away
accounts

Stop using
data harvesters

Protect your privacy

Provide opt-out
options

Use privacy
tools

Use throw-away
accounts

Stop using
data harvesters

Protect your privacy

Provide opt-out
options

Use strong
anonymization

Use privacy
tools

Use throw-away
accounts

Stop using
data harvesters

Protect your privacy

Provide opt-out
options

Use strong
anonymization

Support GDPR
and CCPA

Use privacy
tools

Use throw-away
accounts

Stop using
data harvesters

Protect your privacy

Provide opt-out
options

Use strong
anonymization

Support GDPR
and CCPA




Eliminate bias



Use diverse
datasets


Eliminate bias



Use diverse
datasets

Exclude protected
attributes

Eliminate bias




Use diverse
datasets

Exclude protected
attributes

Use bias-detection
tools

Eliminate bias



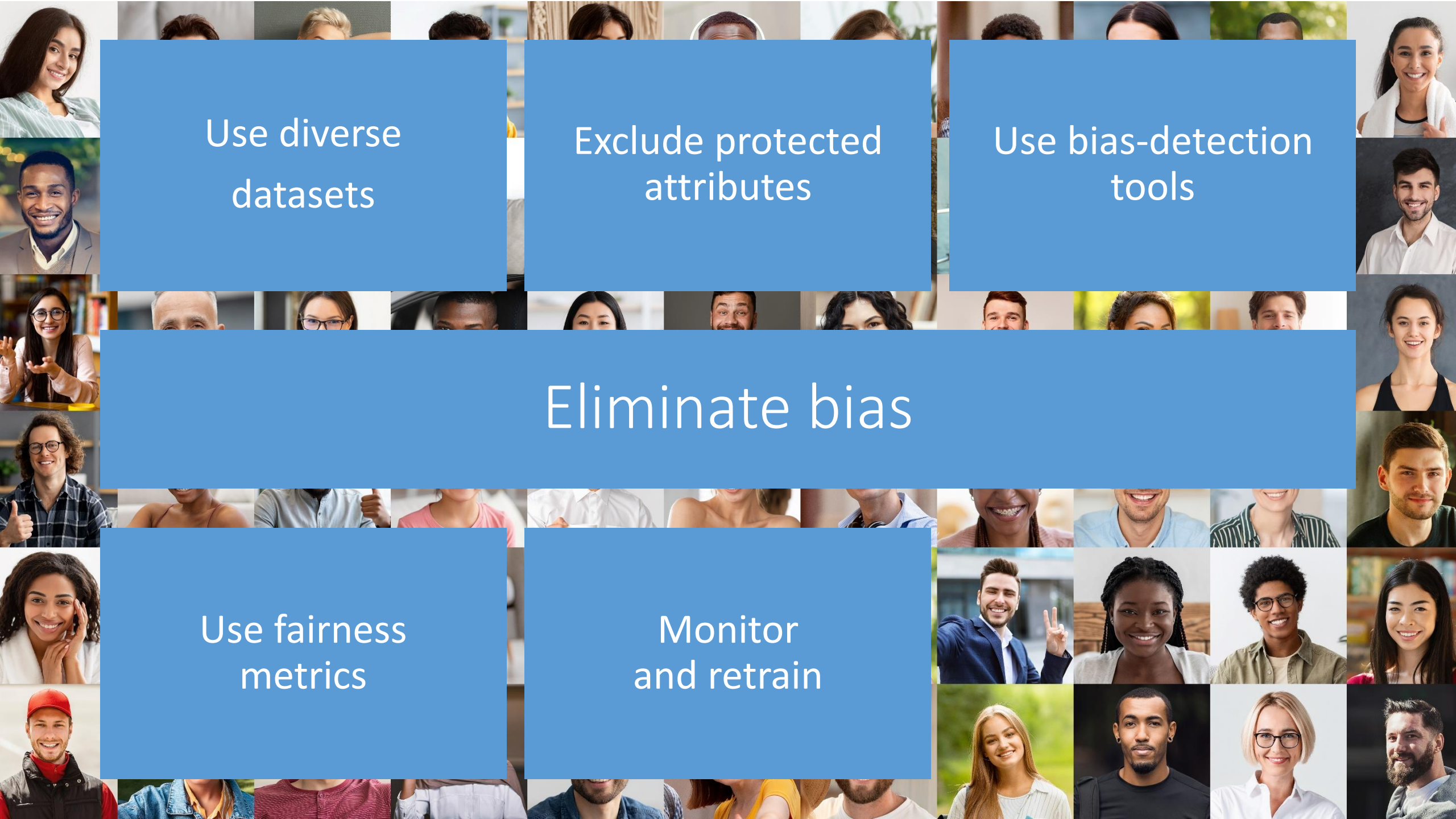
Use diverse
datasets

Exclude protected
attributes

Use bias-detection
tools

Eliminate bias

Use fairness
metrics



Use diverse
datasets

Exclude protected
attributes

Use bias-detection
tools

Eliminate bias

Use fairness
metrics

Monitor
and retrain



Use diverse
datasets

Exclude protected
attributes

Use bias-detection
tools

Eliminate bias

Use fairness
metrics

Monitor
and retrain

Support fairness
regulations

A conceptual image showing a human hand in a grey suit sleeve shaking a white, futuristic robotic hand. The background is a light grey gradient with numerous small, glowing orange and white particles, suggesting a digital or AI environment. A blue horizontal bar is overlaid across the center of the image, containing the text "Create trustworthy AI".

Create trustworthy AI

A hand in a grey suit sleeve is shown holding a white, futuristic robotic hand. The background is a light, hazy space filled with numerous small, glowing white and orange particles, suggesting a digital or futuristic environment. A blue rectangular box is positioned in the upper left, and a wider blue horizontal bar spans across the middle of the image.

Choose
transparency

Create trustworthy AI

A conceptual image showing a human hand in a grey suit jacket shaking a white, futuristic robotic hand. The background is light grey with faint white lines and small orange and white particles. Two blue rectangular boxes are positioned at the top, and a larger blue horizontal bar is in the middle.

Choose
transparency

Keep it
simple

Create trustworthy AI



Choose
transparency

Keep it
simple

Use explainable
AI tools

Create trustworthy AI



Choose
transparency

Keep it
simple

Use explainable
AI tools

Create trustworthy AI

Provide uncertainty
measures



Choose
transparency

Keep it
simple

Use explainable
AI tools

Create trustworthy AI

Provide uncertainty
measures

Keep a human
in the loop

Choose
transparency

Keep it
simple

Use explainable
AI tools

Create trustworthy AI

Provide uncertainty
measures

Keep a human
in the loop

Provide
recourse

A magnifying glass with a black handle and frame is positioned over an open book. The book's pages are white and slightly blurred. In the foreground, a silver and black pen lies horizontally across the bottom of the frame. The background is a plain, light-colored surface.

Fight misinformation

The background of the slide features a blurred image of an open book with a magnifying glass resting on it. A pen is also visible in the foreground, lying horizontally across the bottom of the frame. The overall scene is set against a light, neutral background.

Who is the
author?

Fight misinformation

The background of the slide features a blurred image of an open book with a pen resting on its pages. The text is overlaid on blue rectangular boxes.

Who is the
author?

Who is the
publisher?

Fight misinformation

Who is the
author?

Who is the
publisher?

What are
their claims?

Fight misinformation



Who is the
author?

Who is the
publisher?

What are
their claims?

Fight misinformation

How strong is
their evidence?

Who is the author?

Who is the publisher?

What are their claims?

Fight misinformation

How strong is their evidence?

What are their sources?

Who is the author?

Who is the publisher?

What are their claims?

Fight misinformation

How strong is their evidence?

What are their sources?

Be skeptical but open-minded

Social

Get off social media

Cancel

Snapchat

Flickr

Instagram



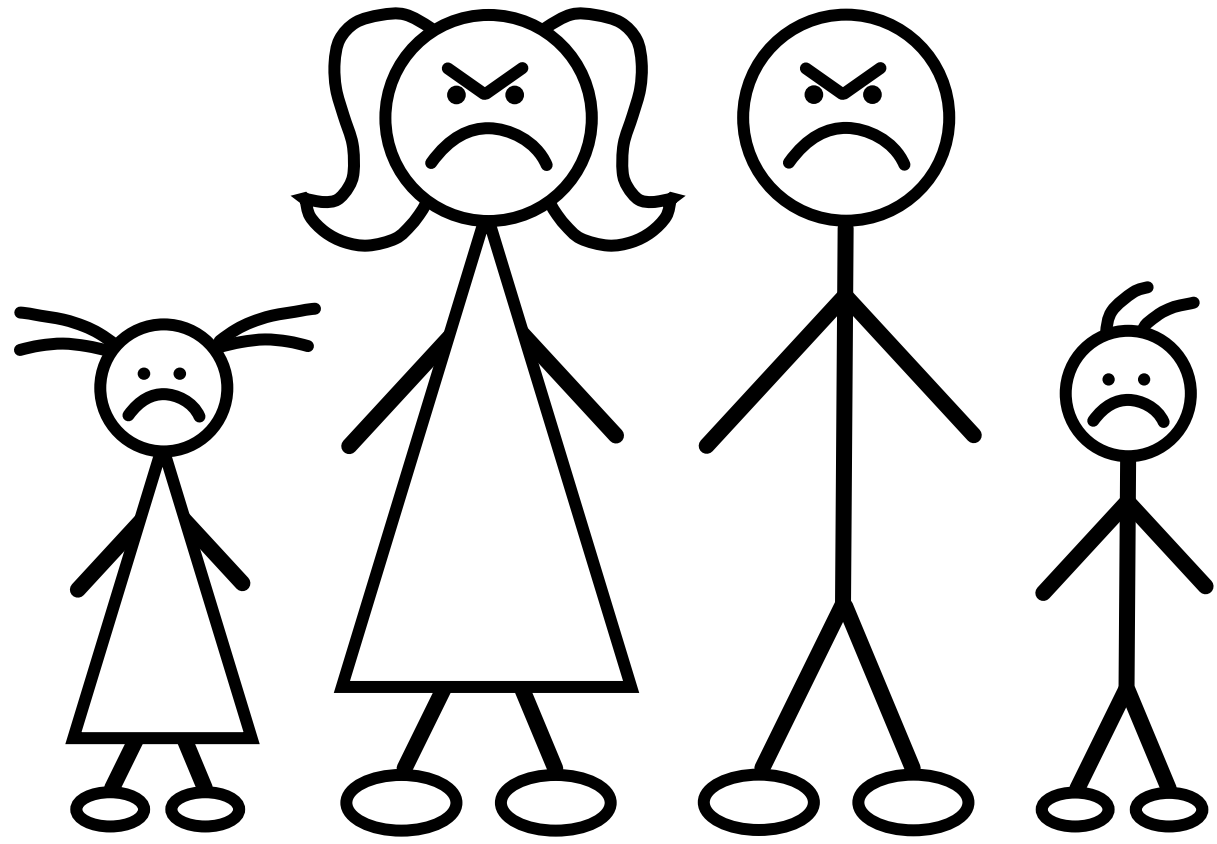
Talk to other people

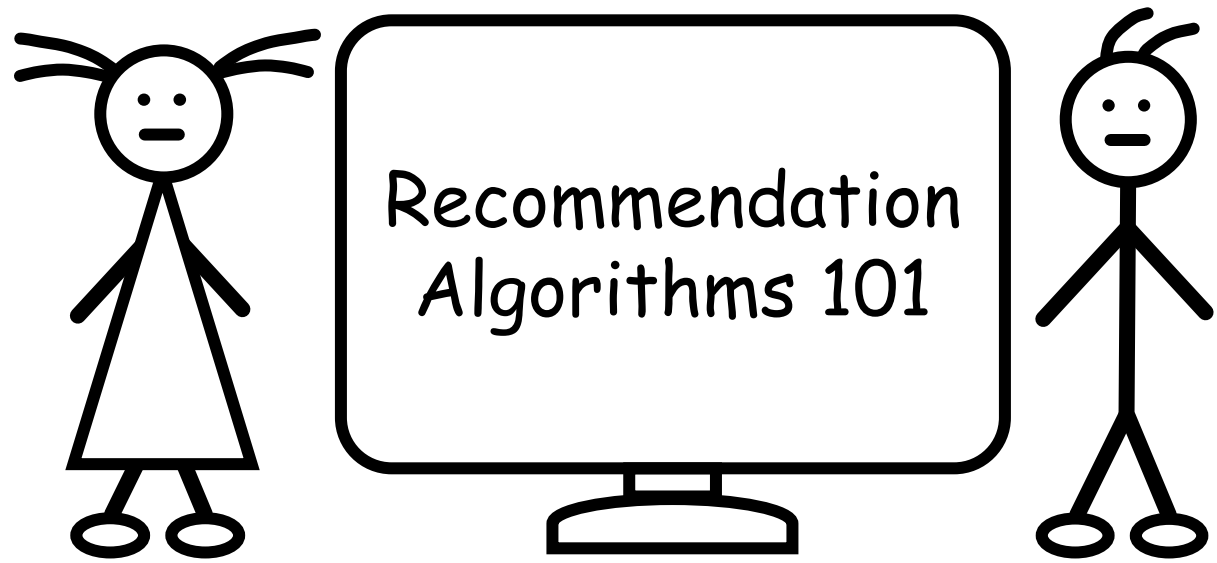


Talk to other people

Alice and Bob in AI Land

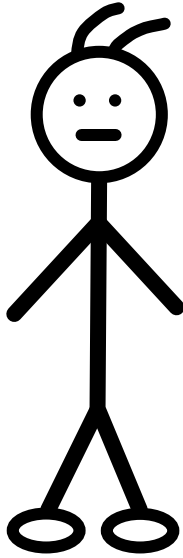
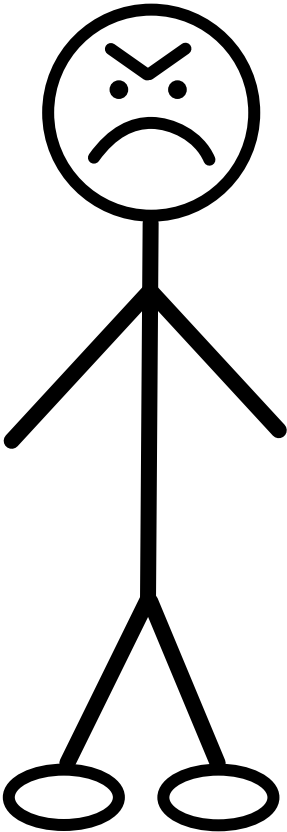
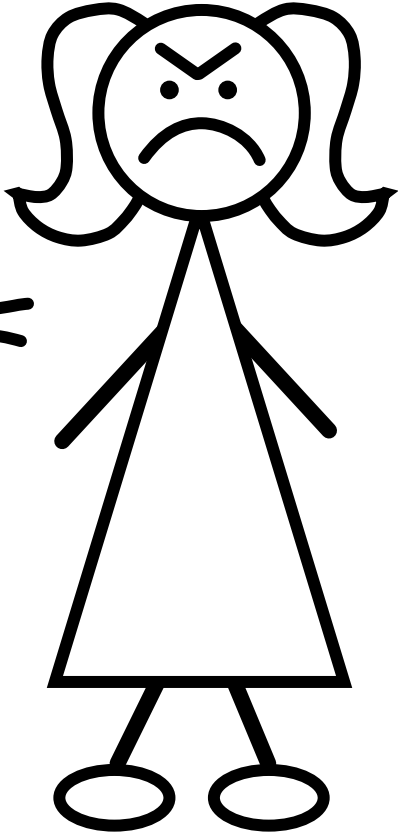
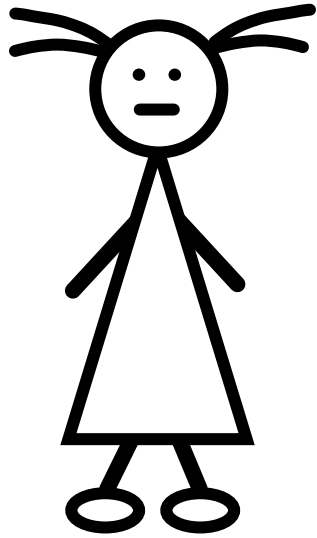
(continued ...)

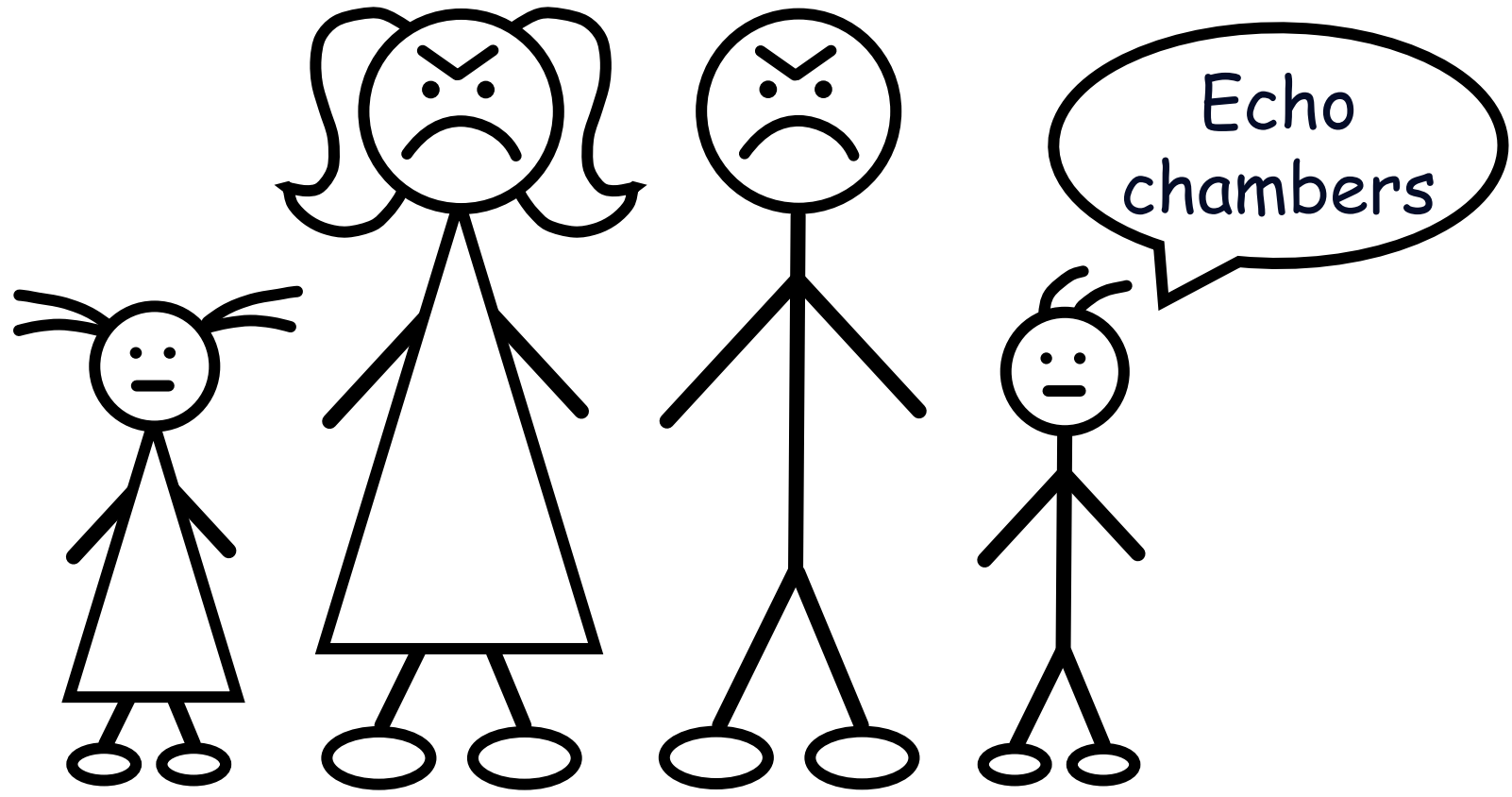




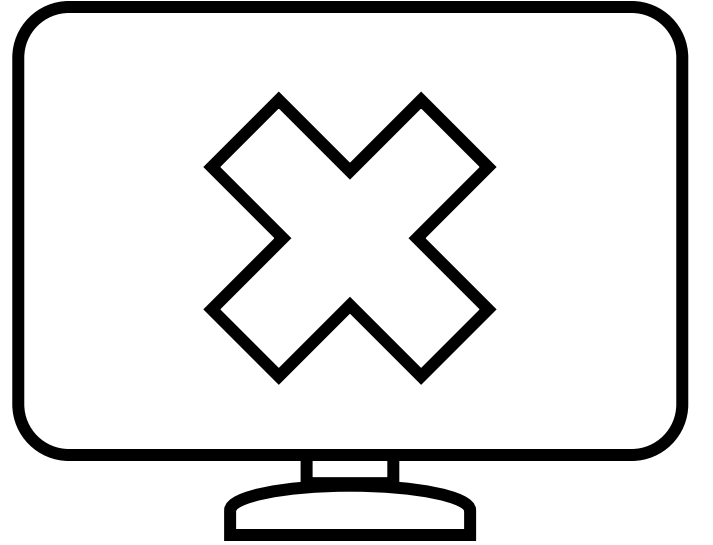
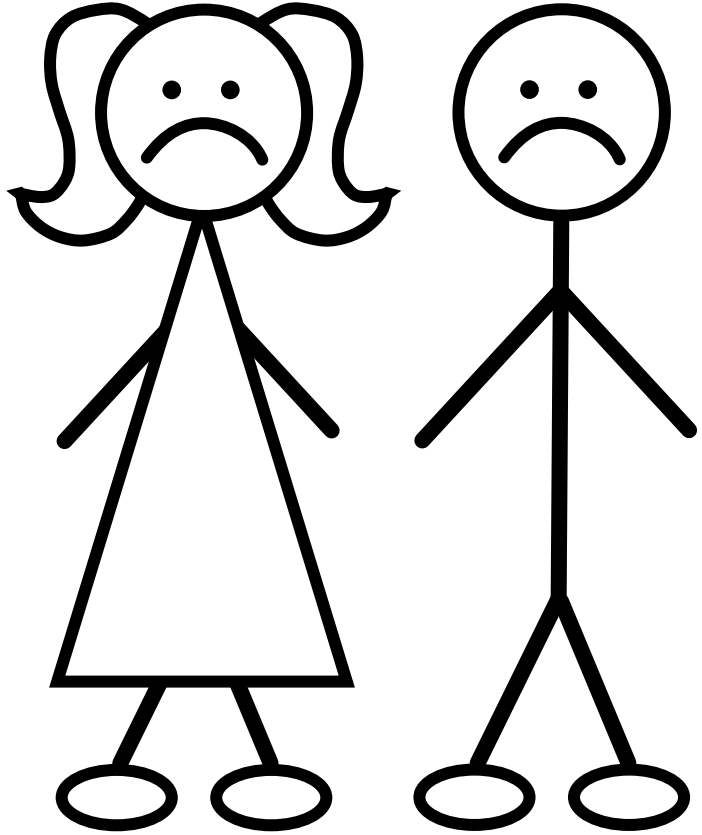
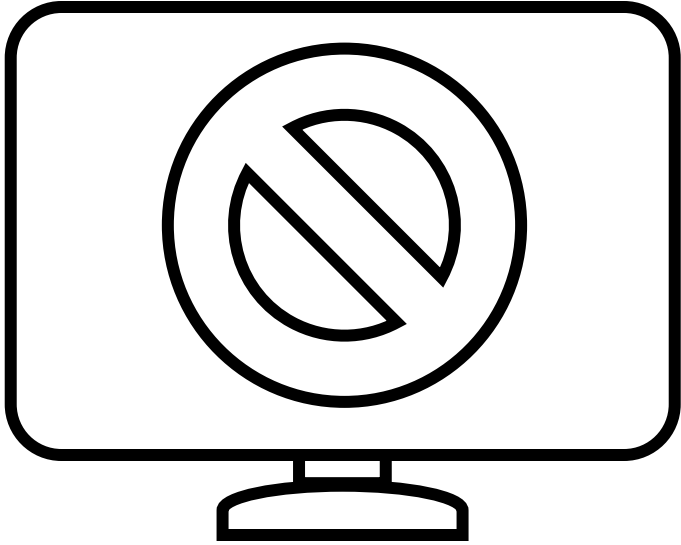
Recommendation
Algorithms 101

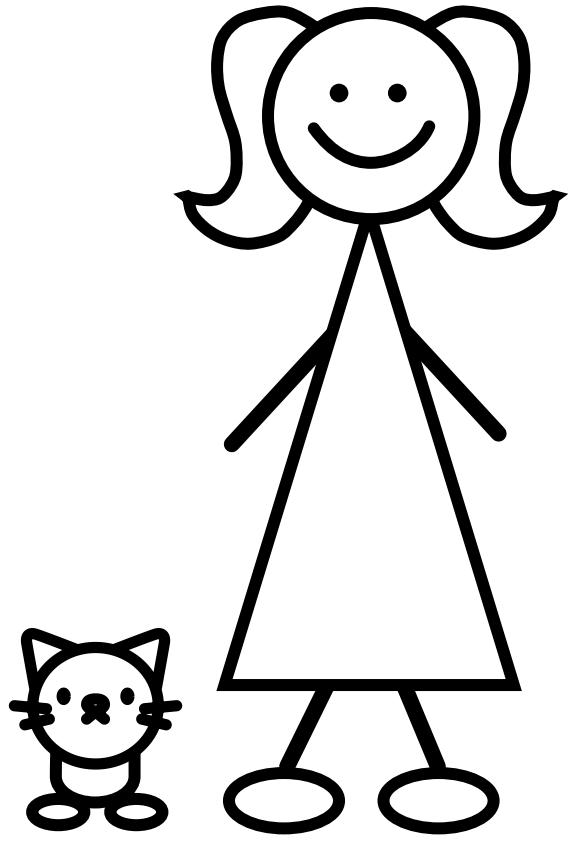
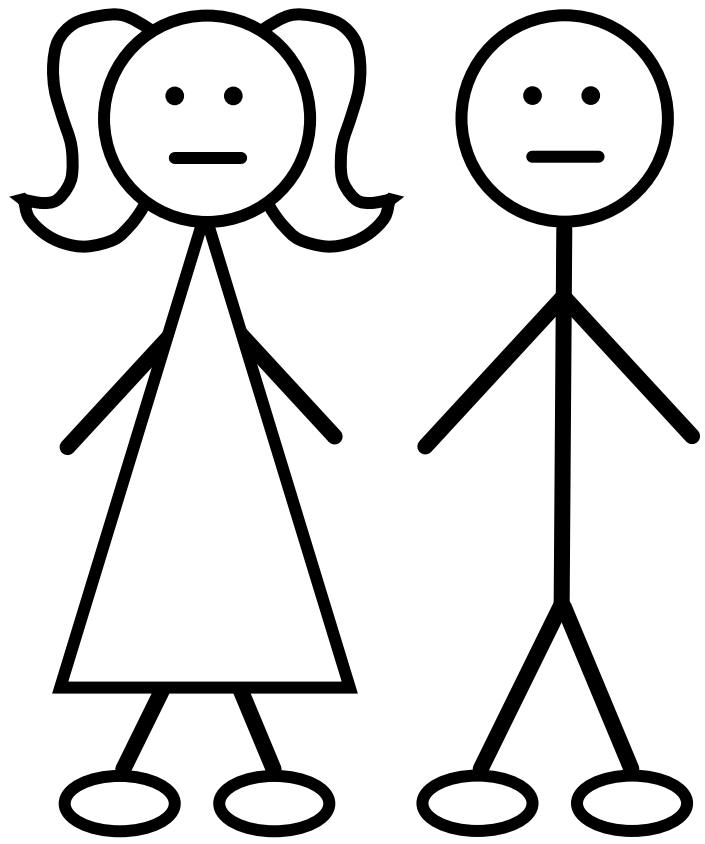
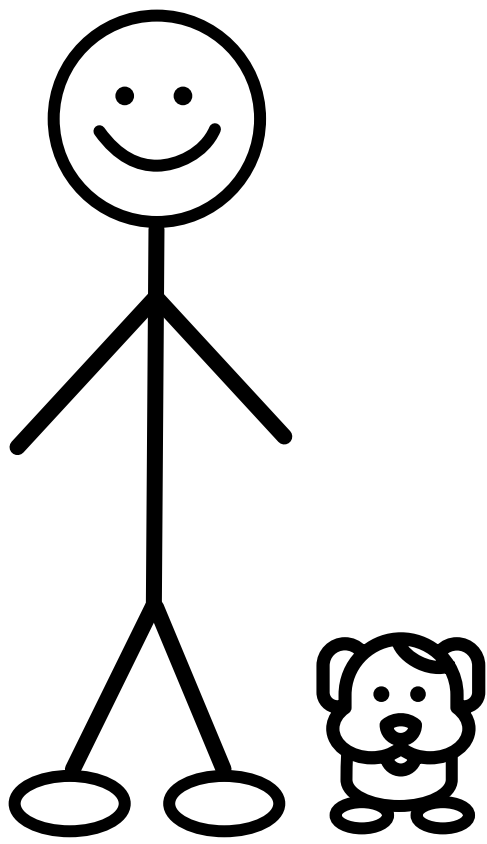
Rabbit
holes

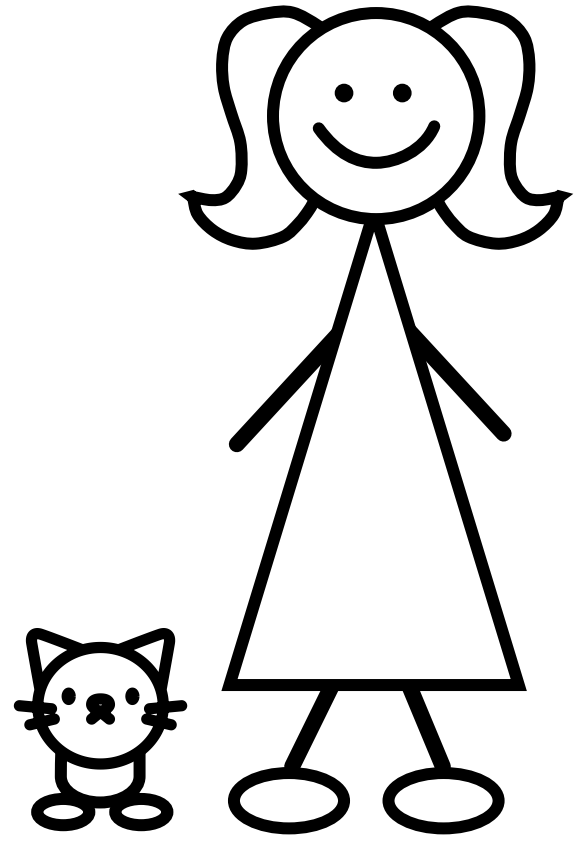
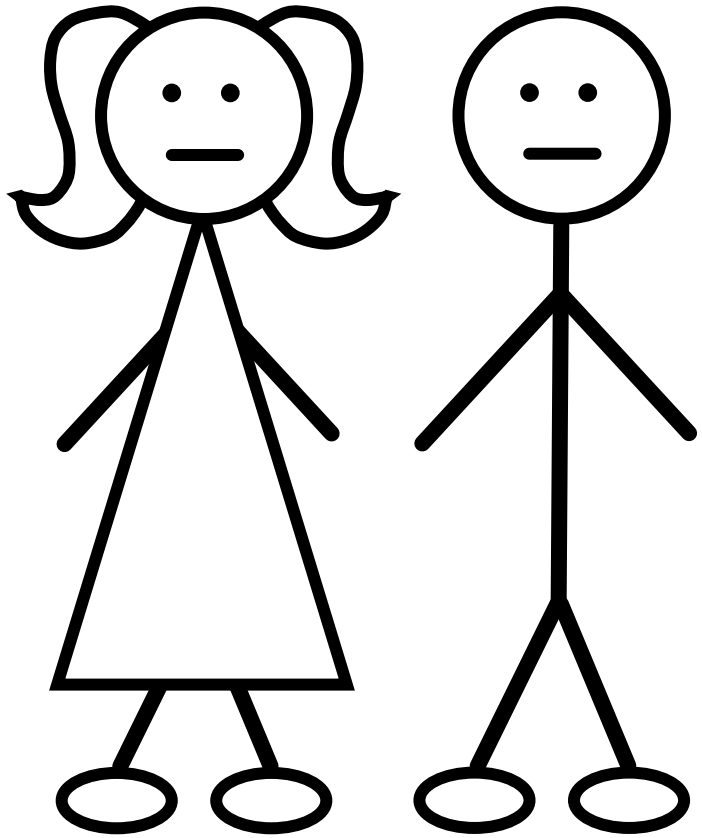
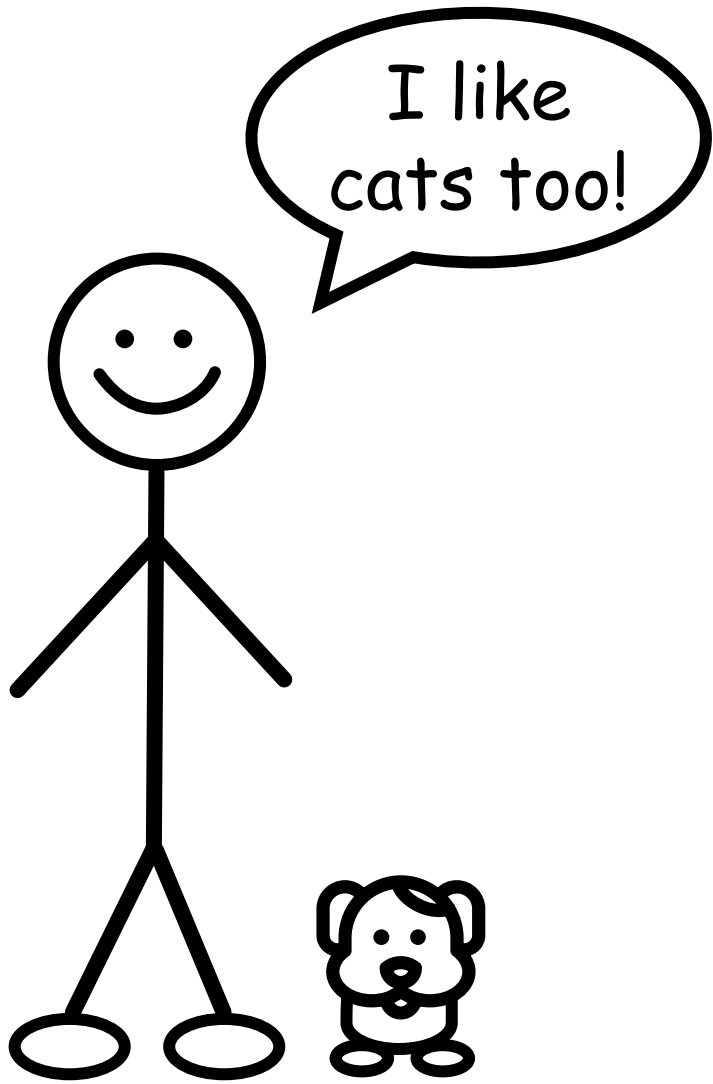


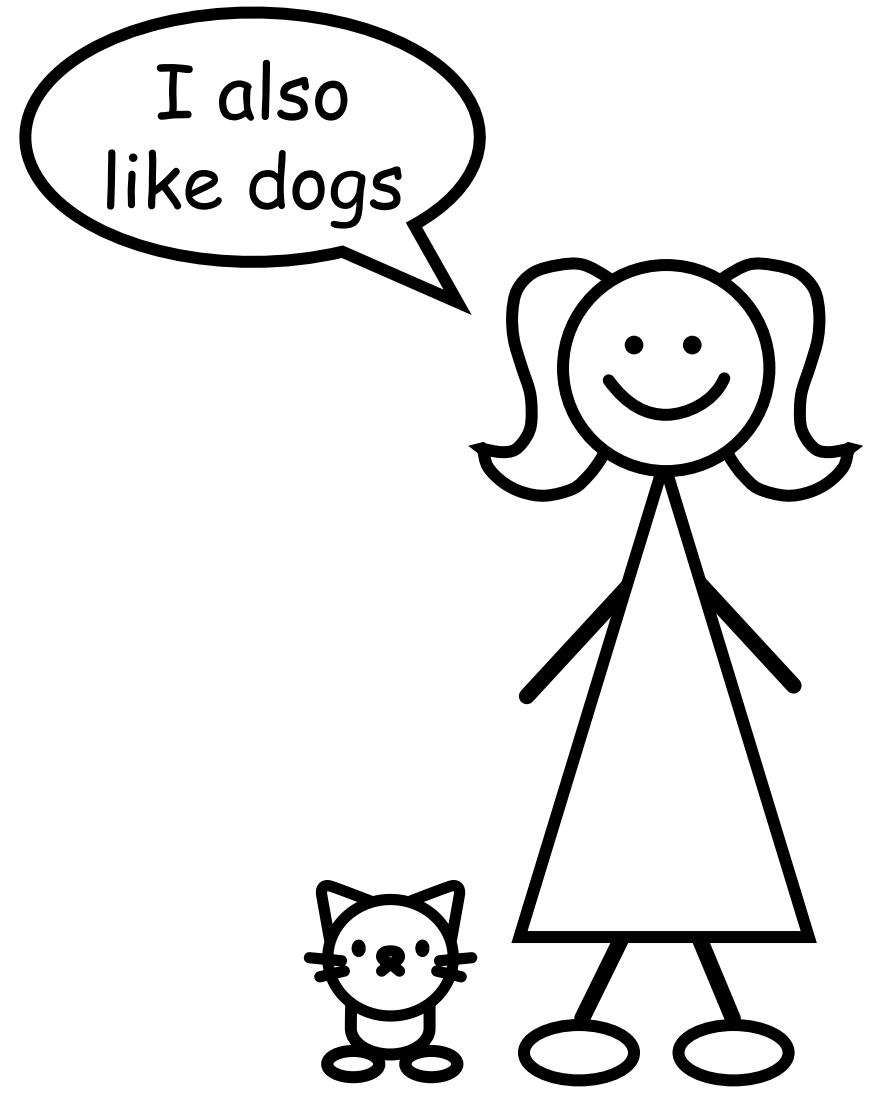
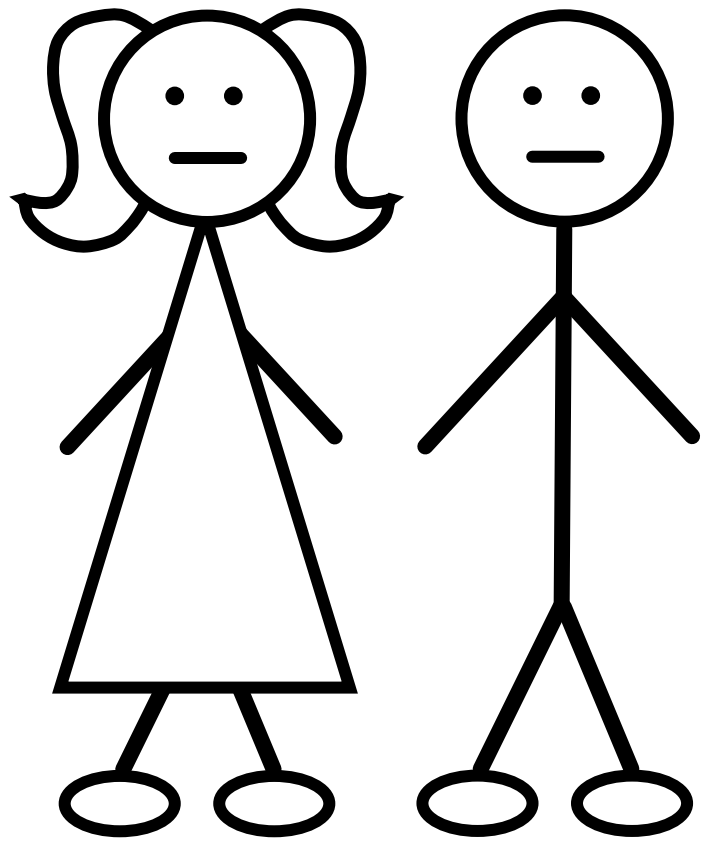
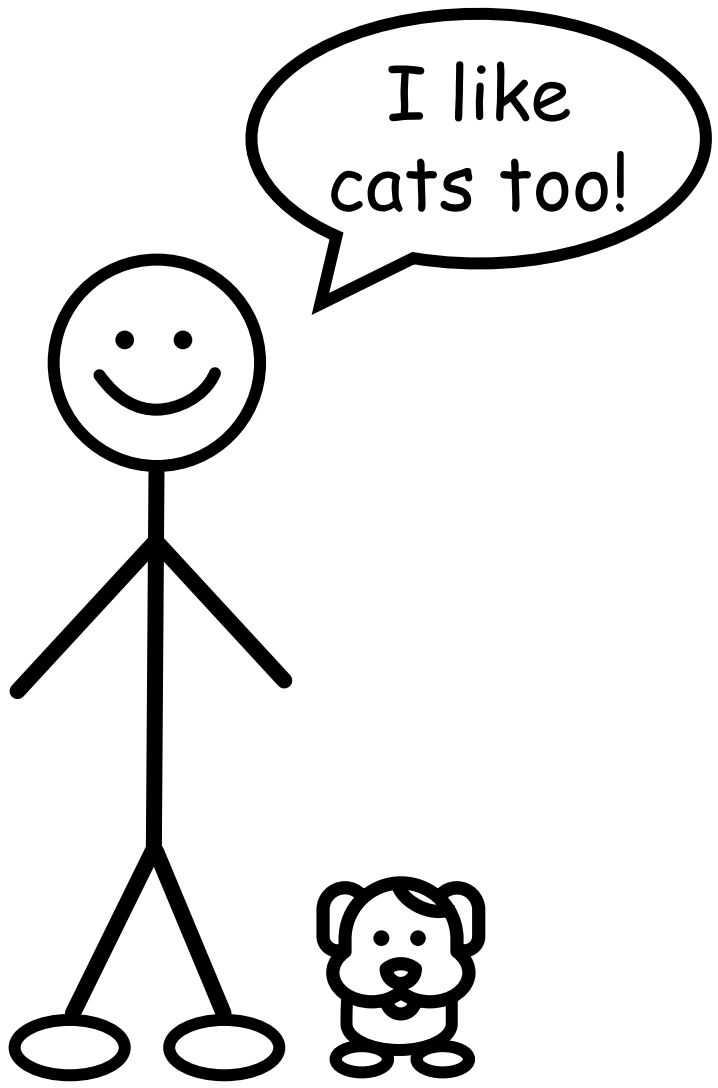


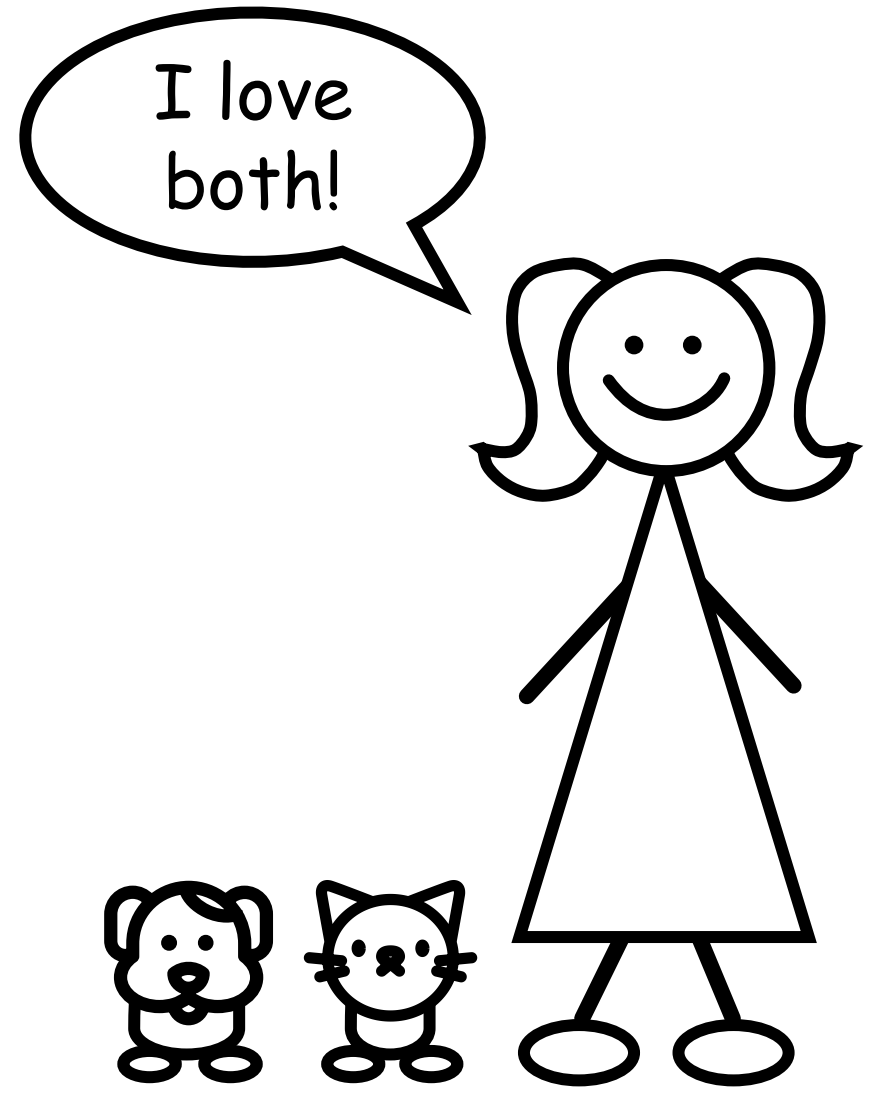
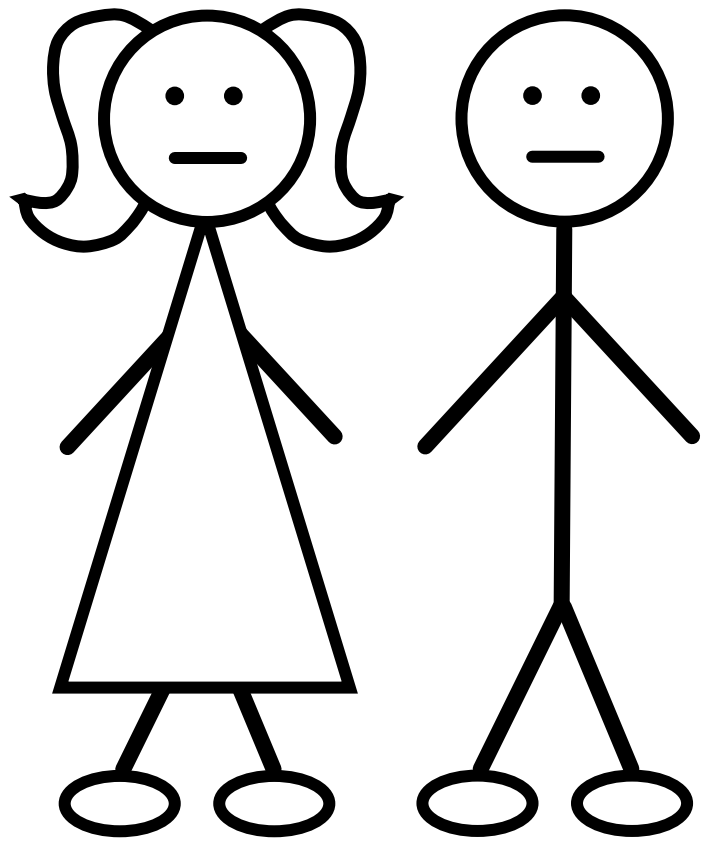
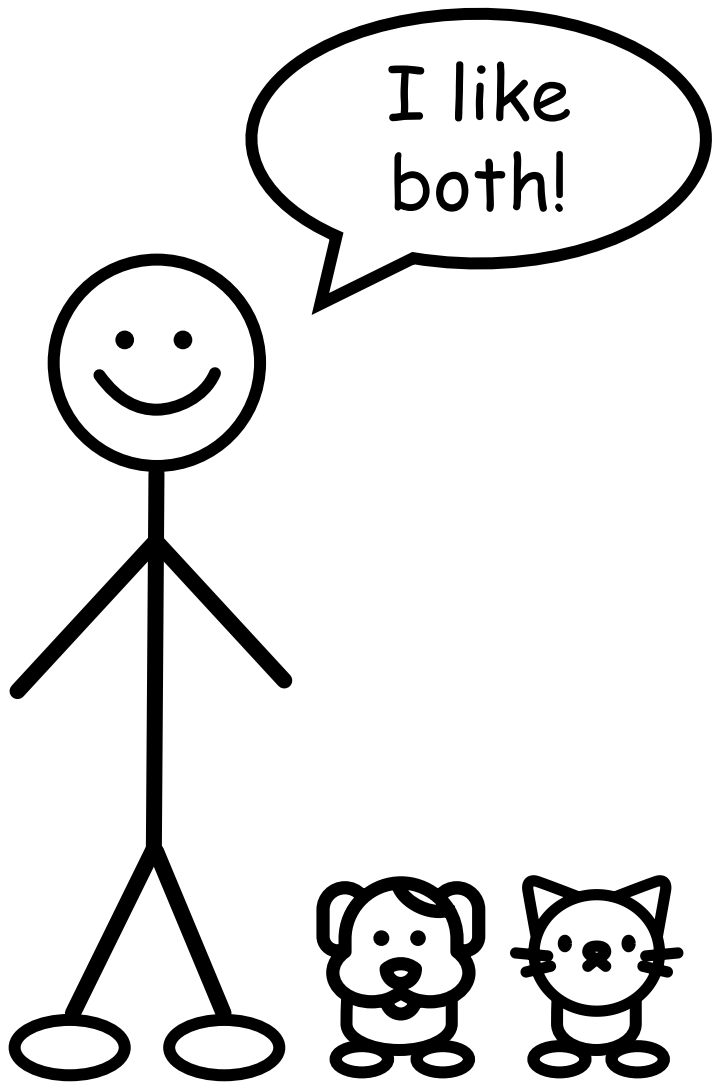
Echo chambers

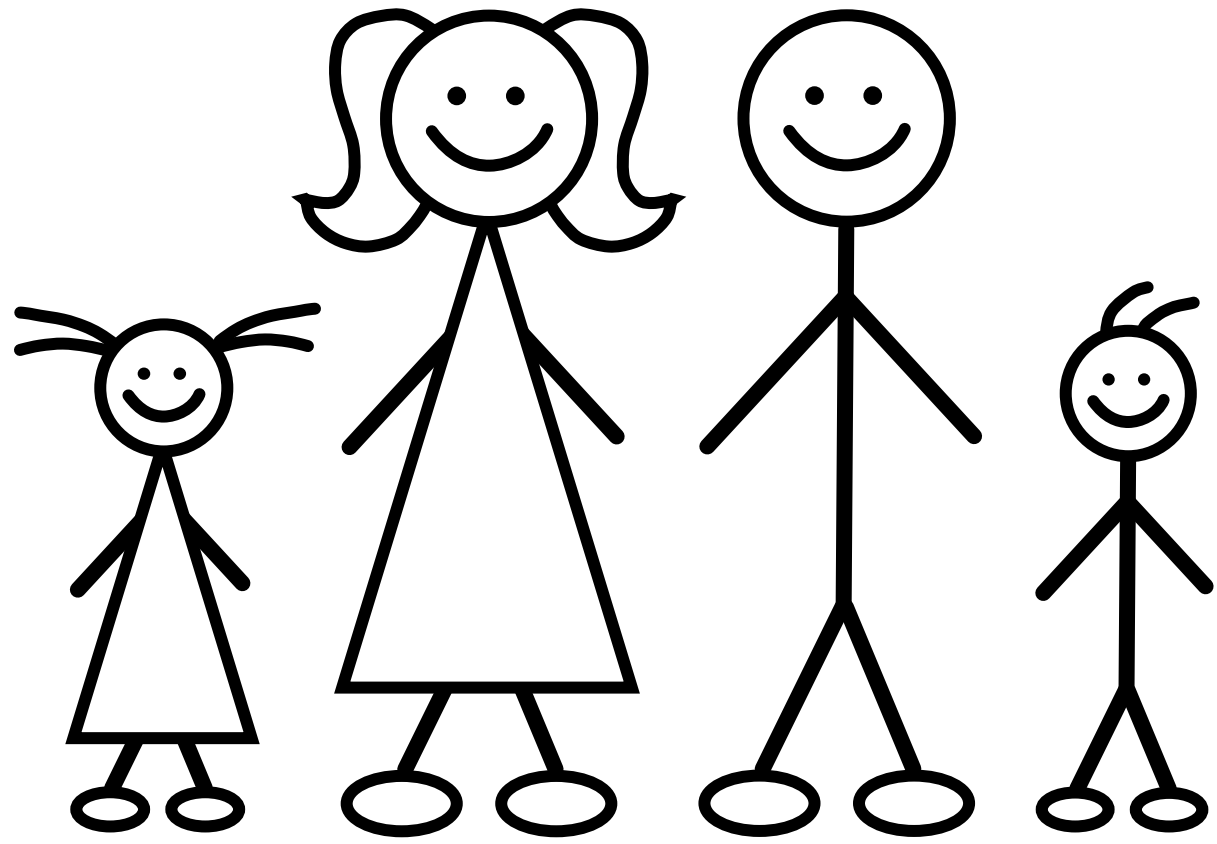


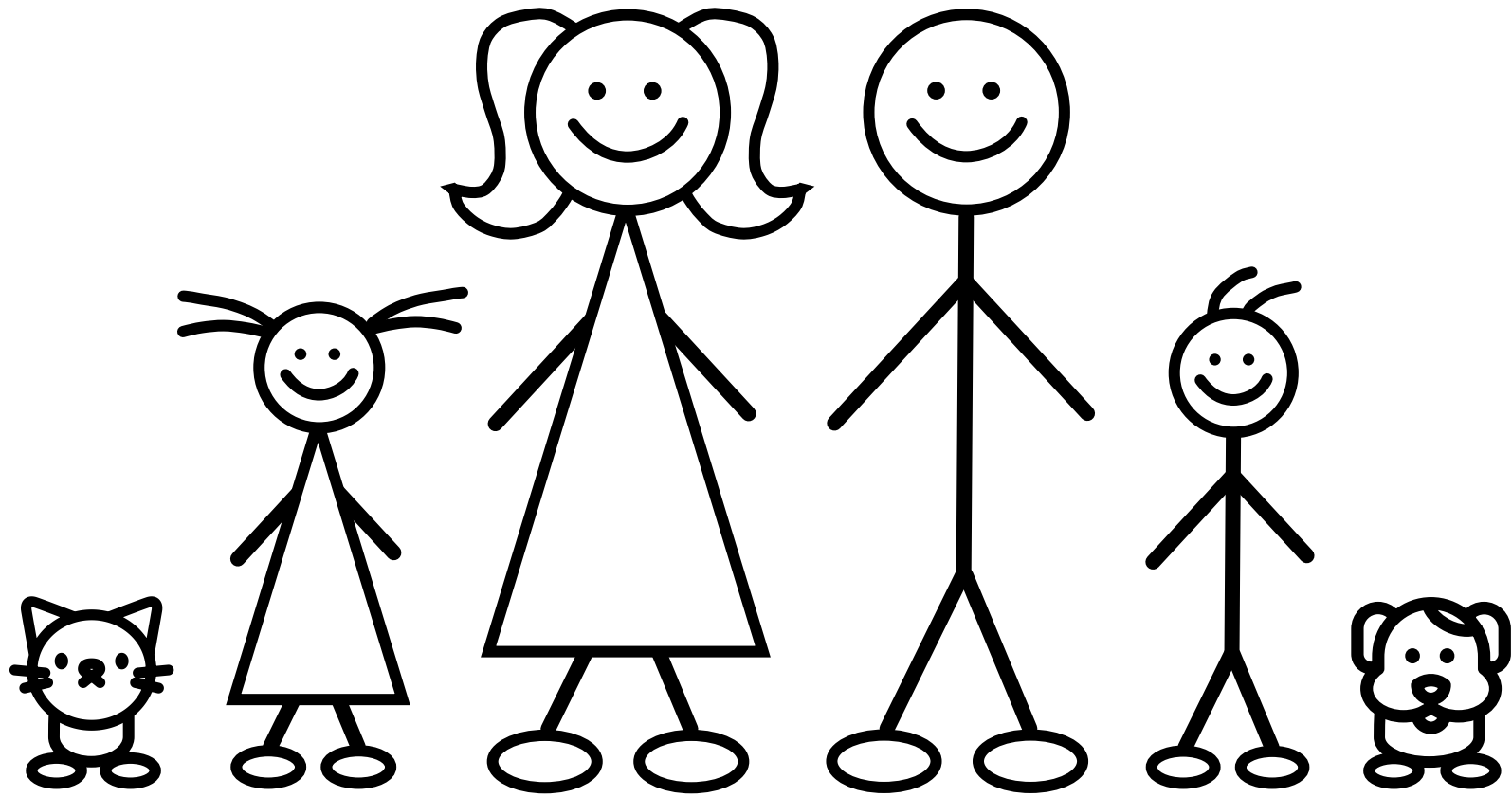


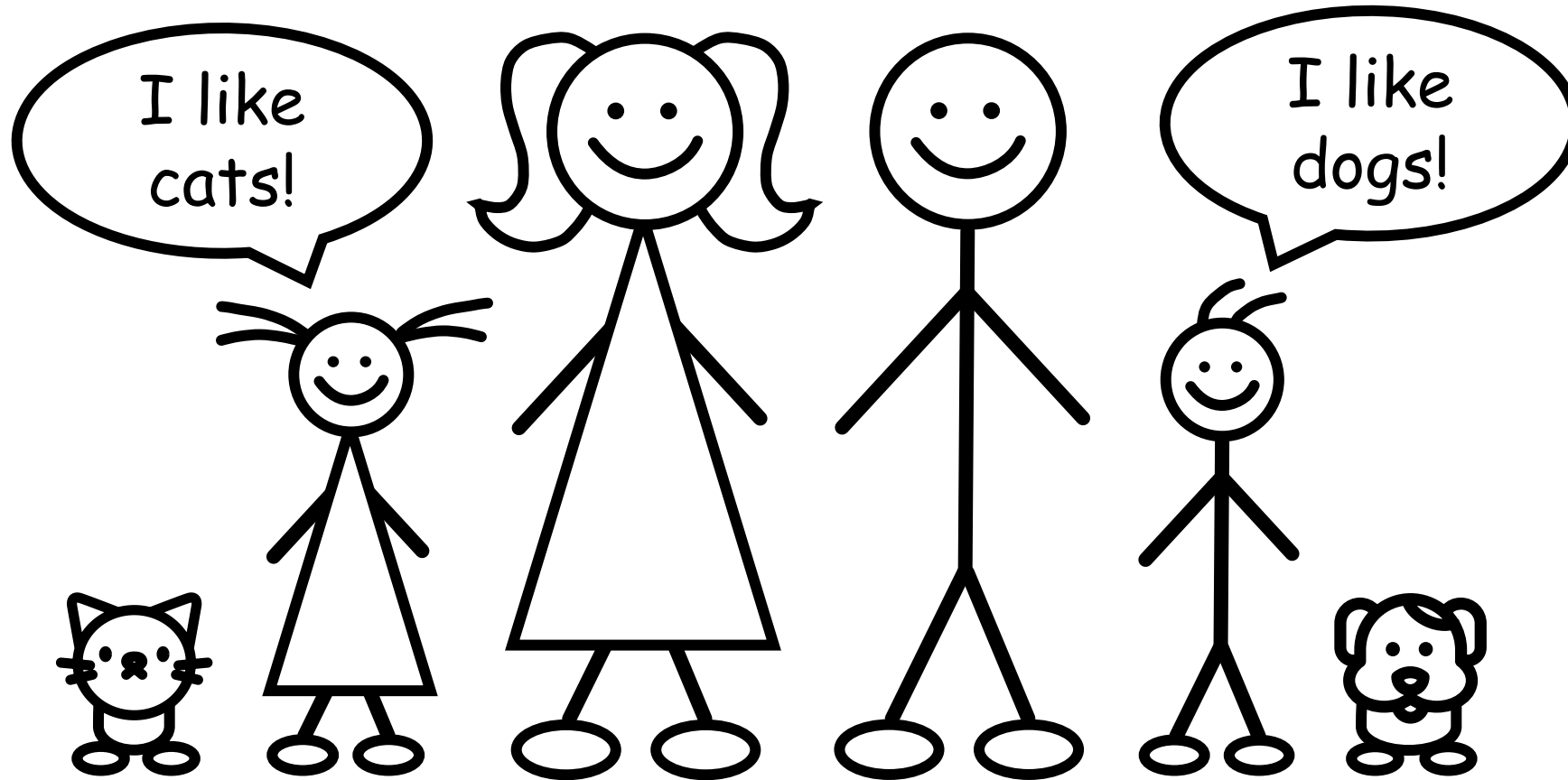








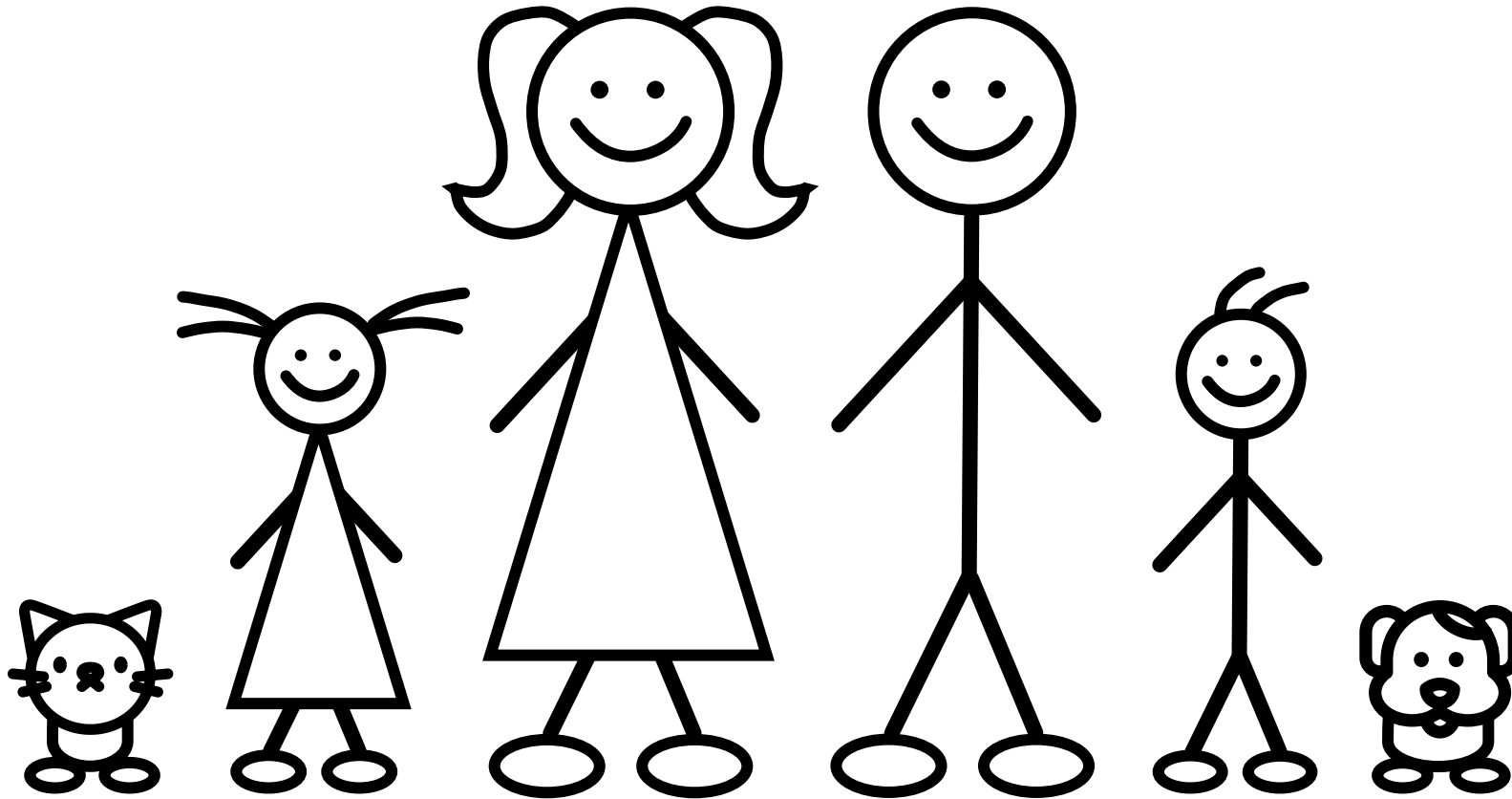




I like
cats!

I like
dogs!

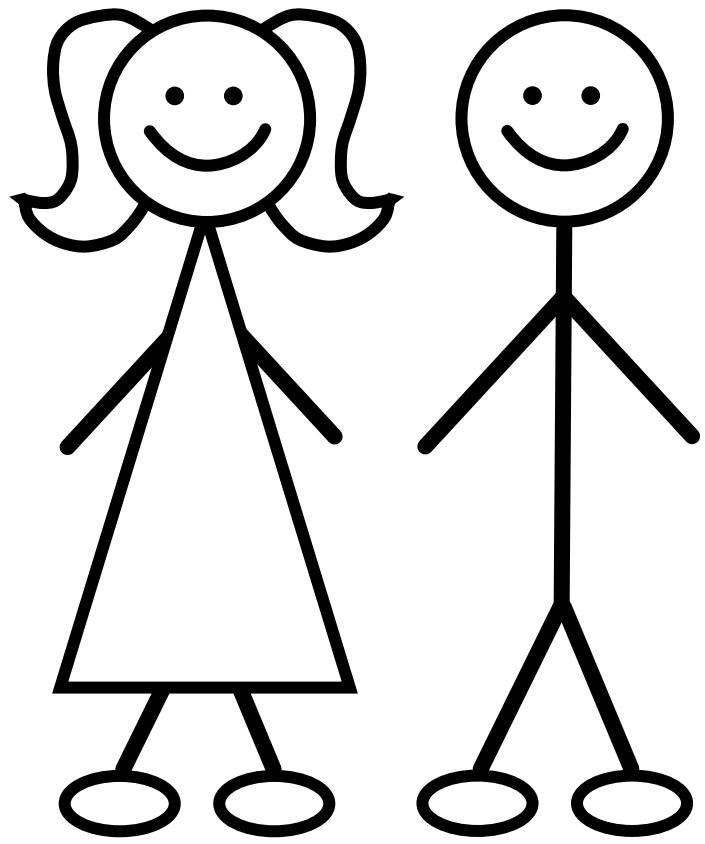
Happily ever after...

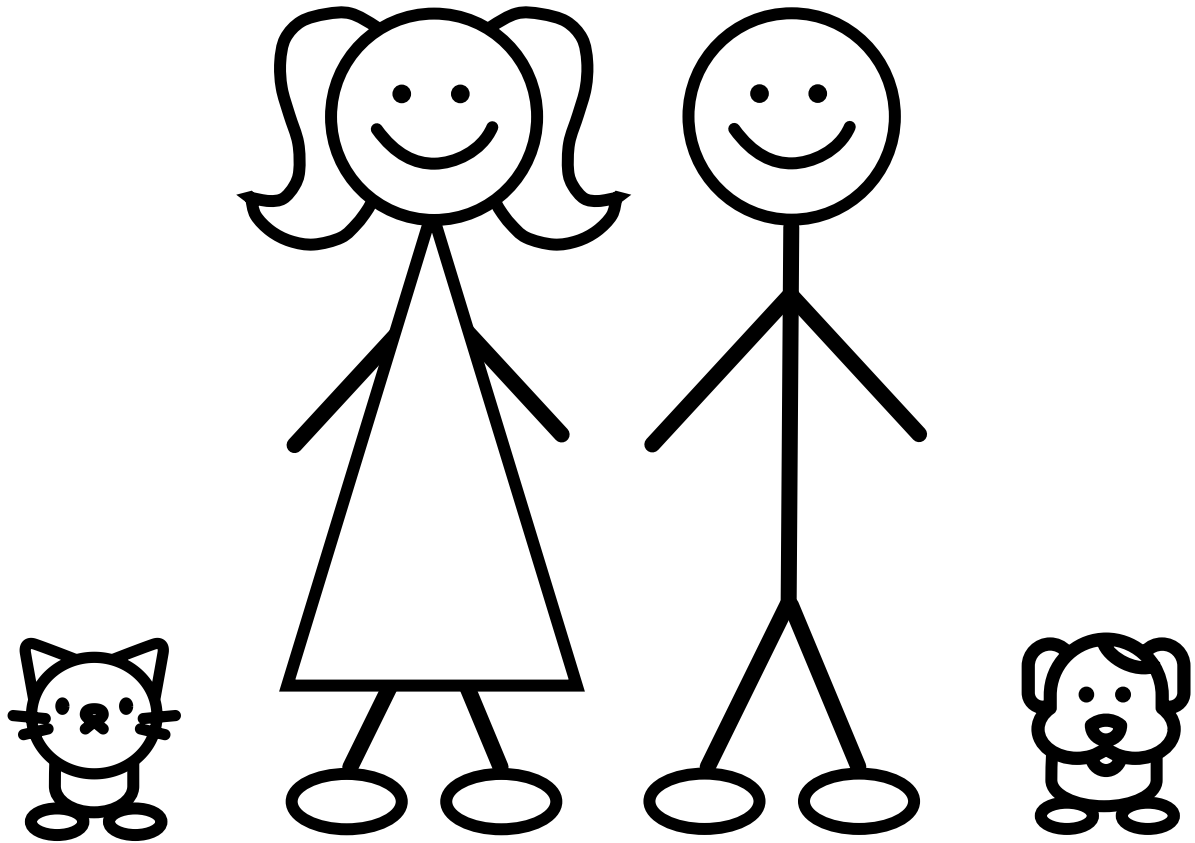


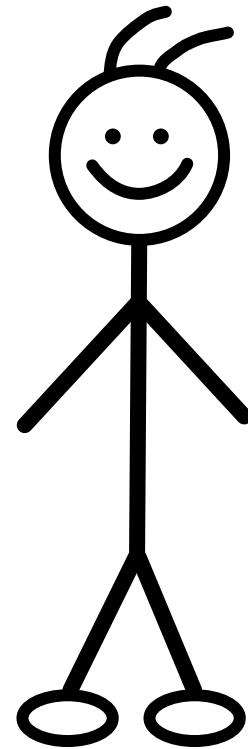
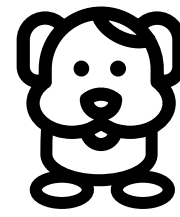
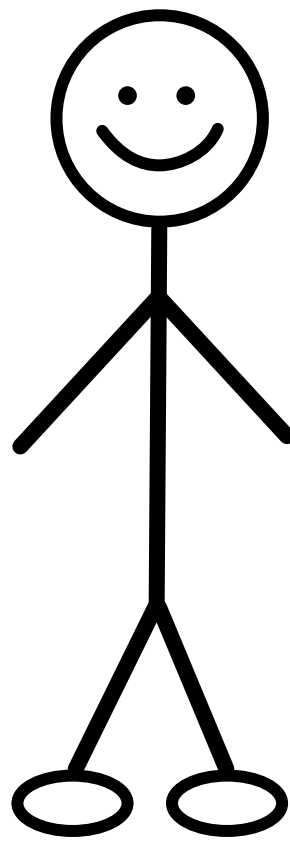
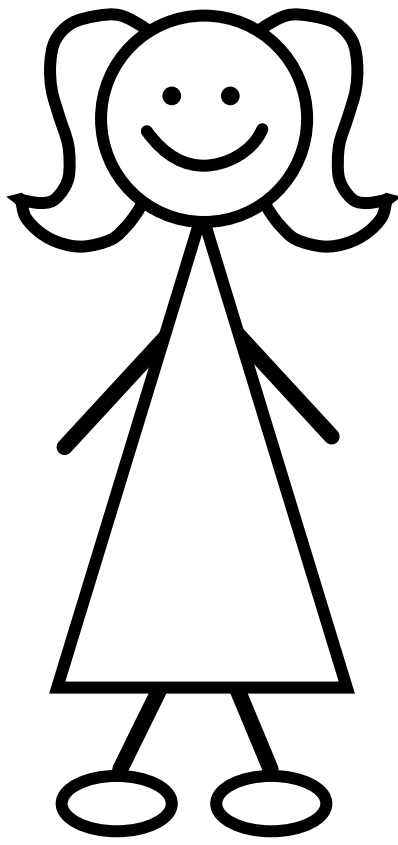
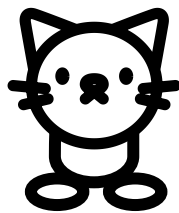
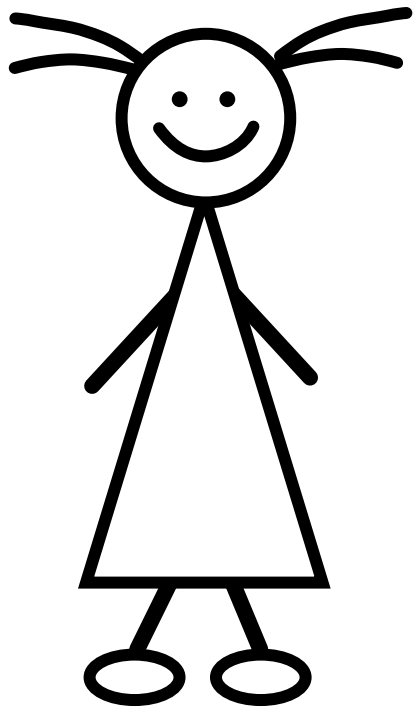
AI Ethics for Tomorrow

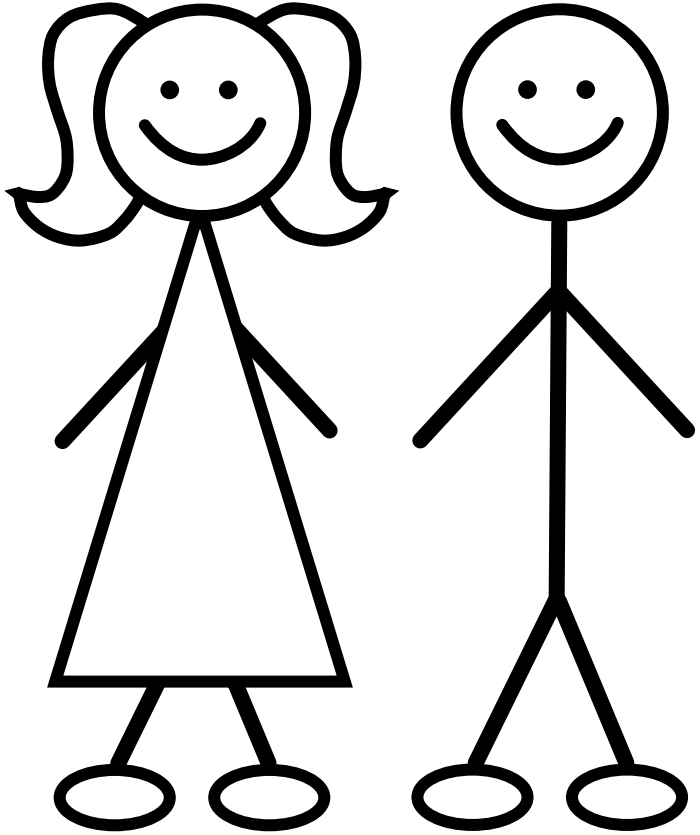
Alice and Bob in AI Land

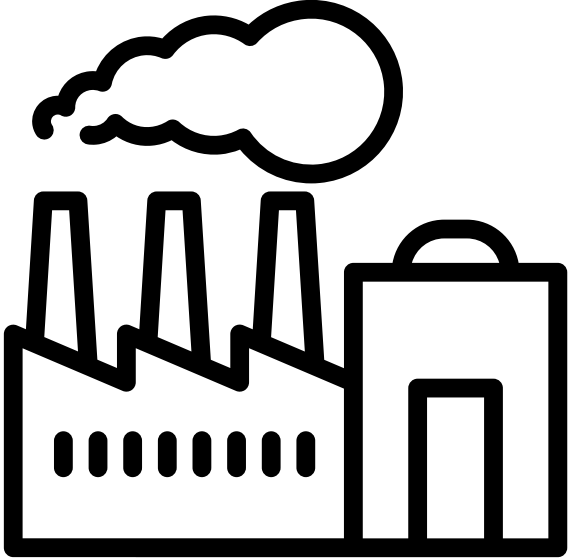
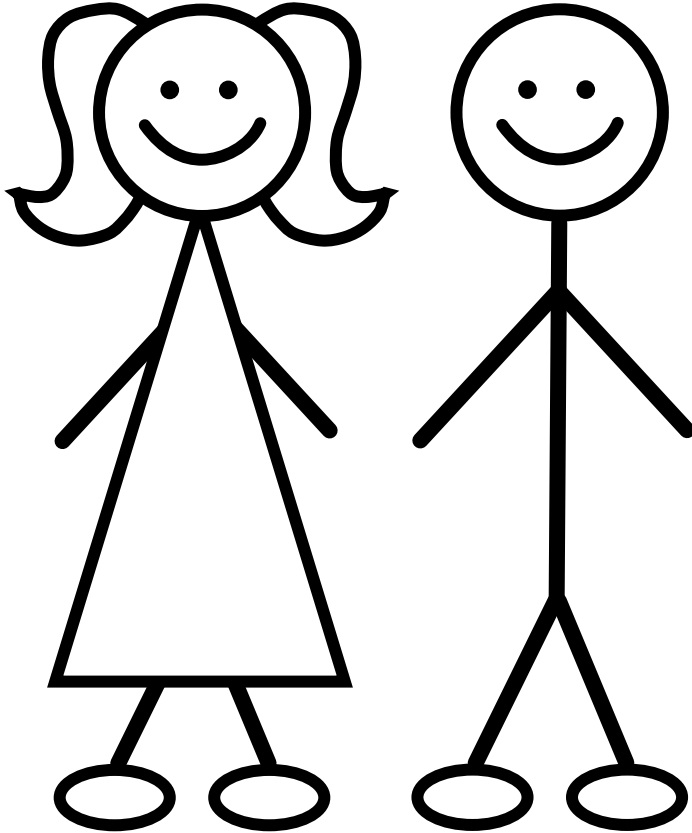
Chapter 2:
"AI Ethics of Tomorrow"

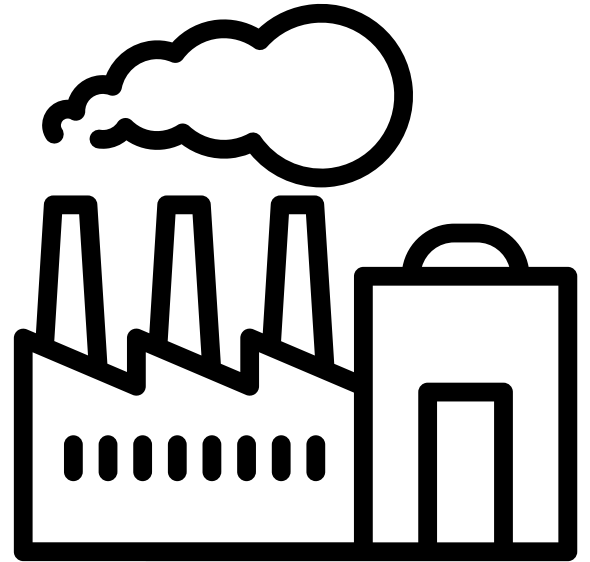
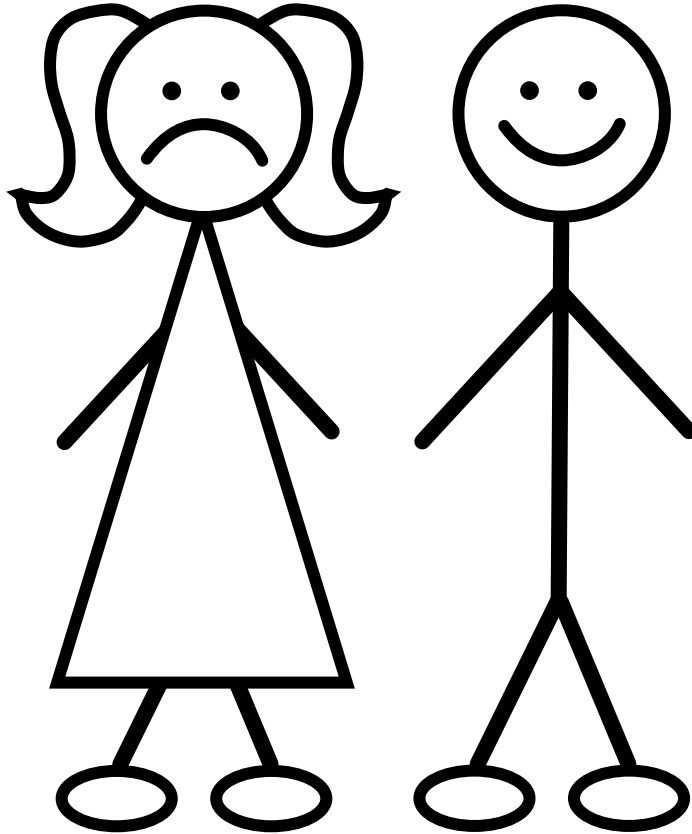
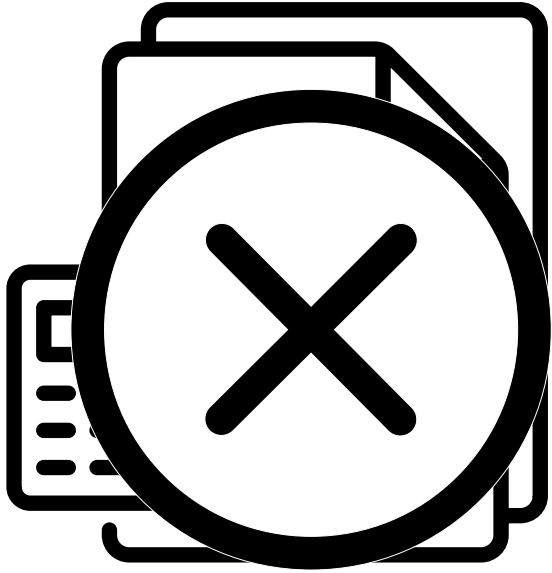


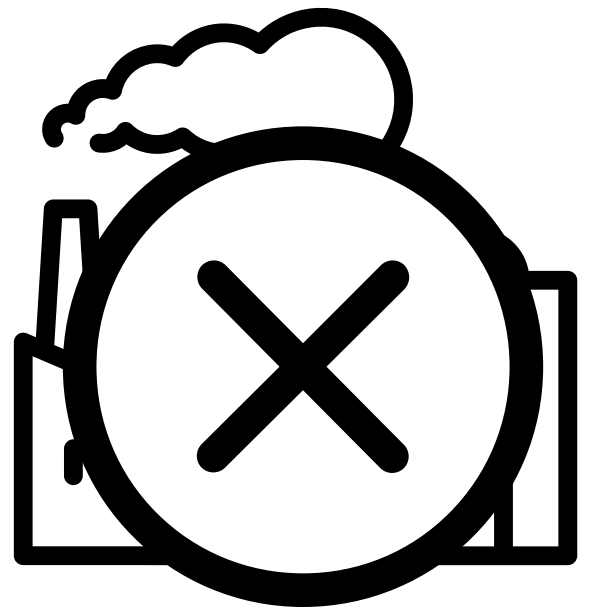
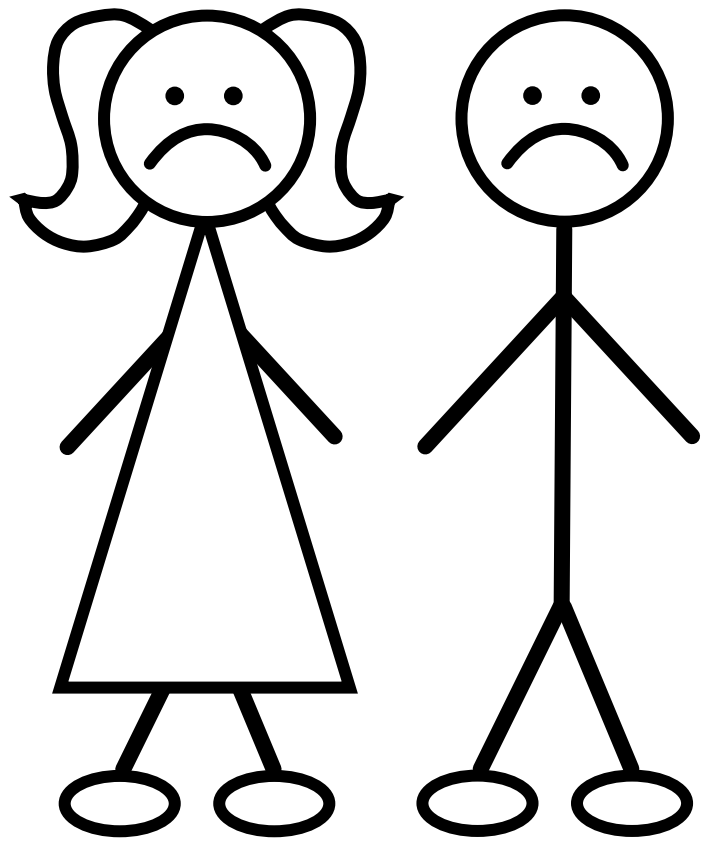


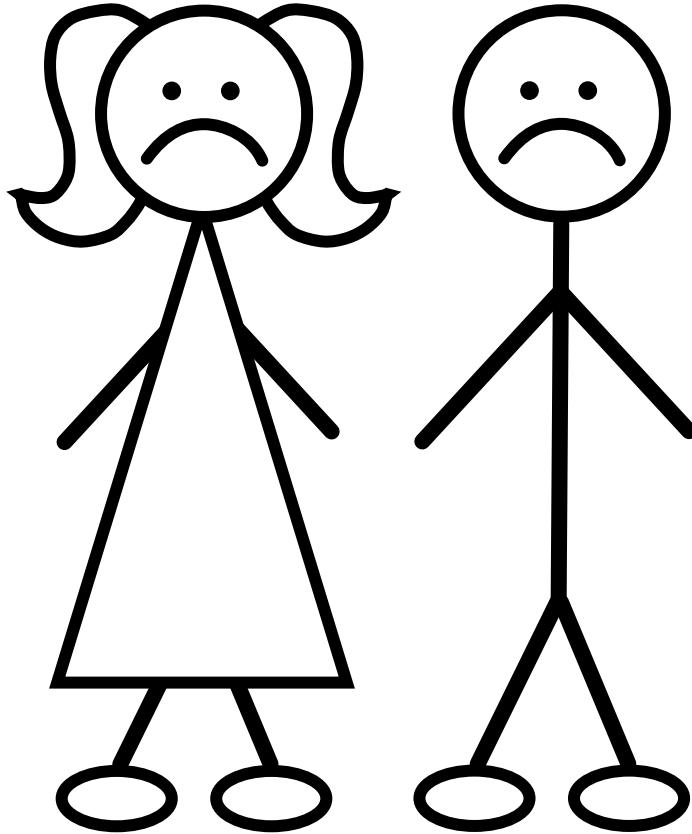


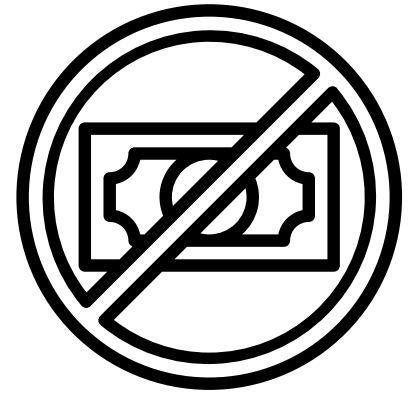
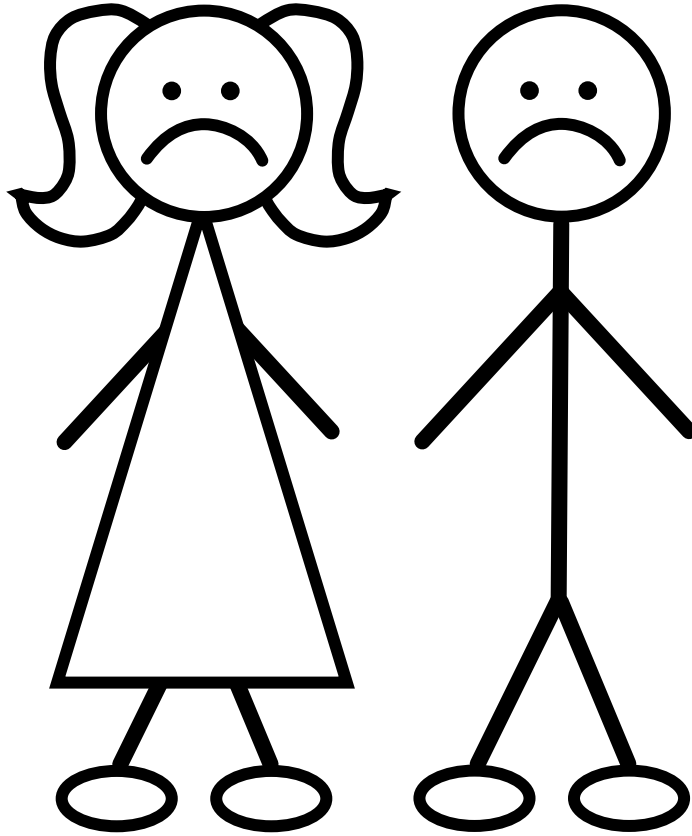


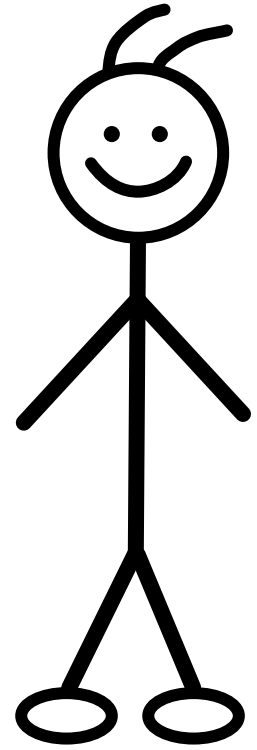
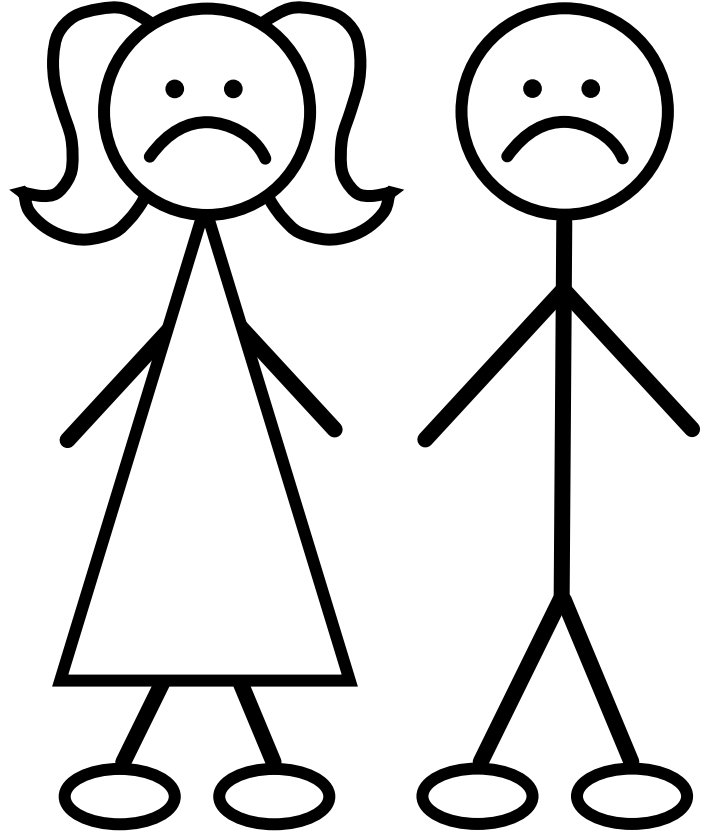
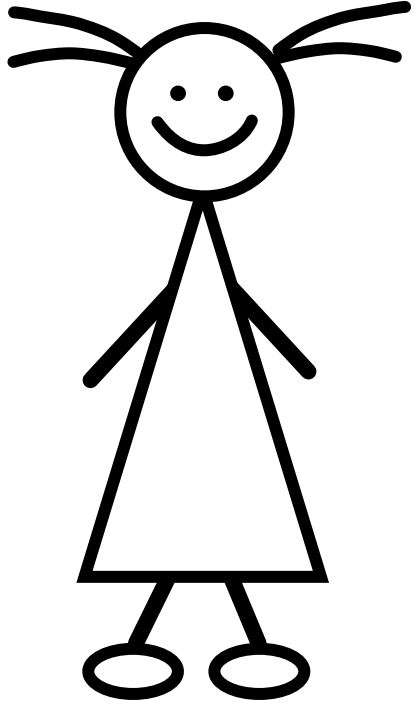




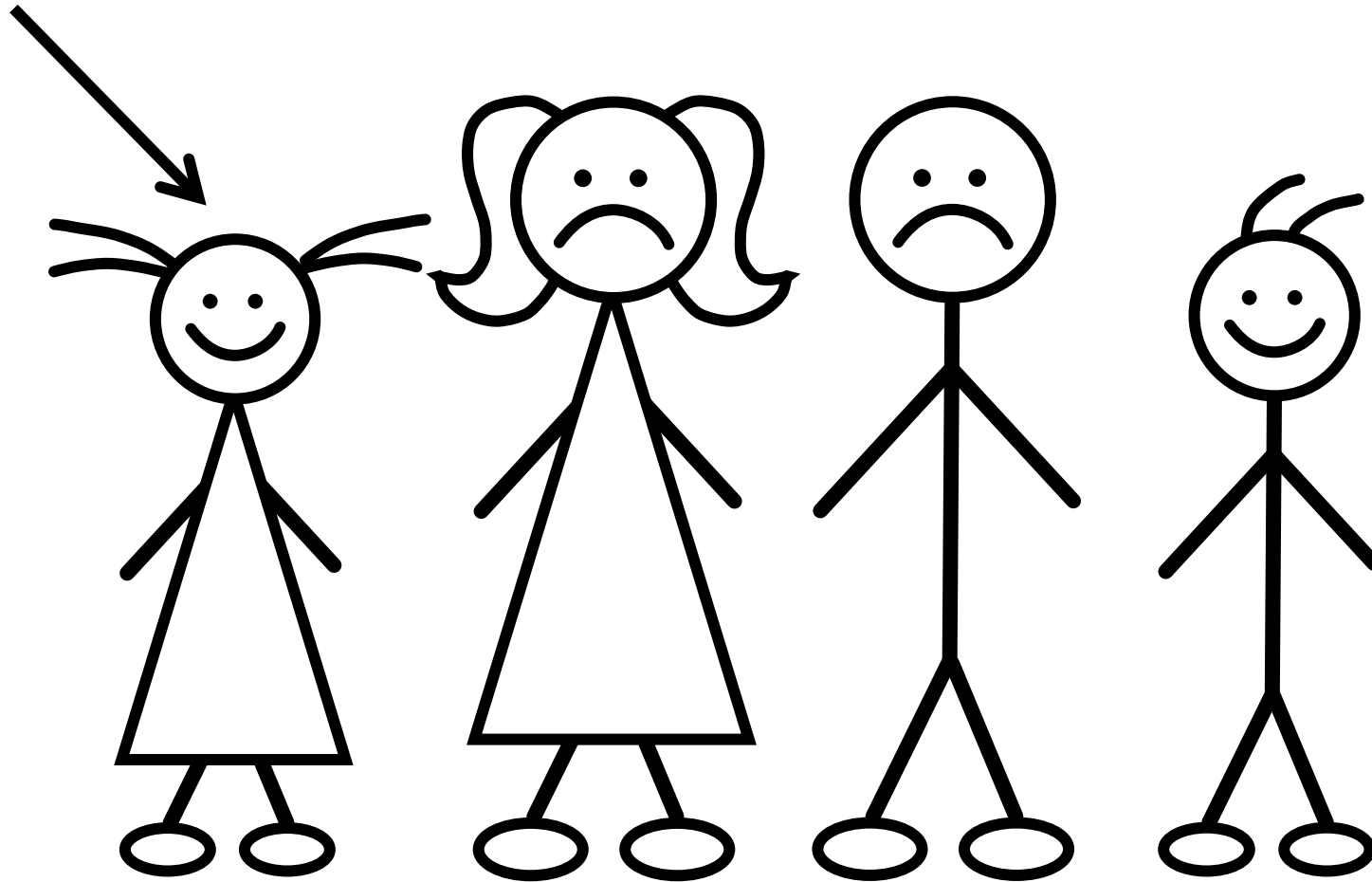




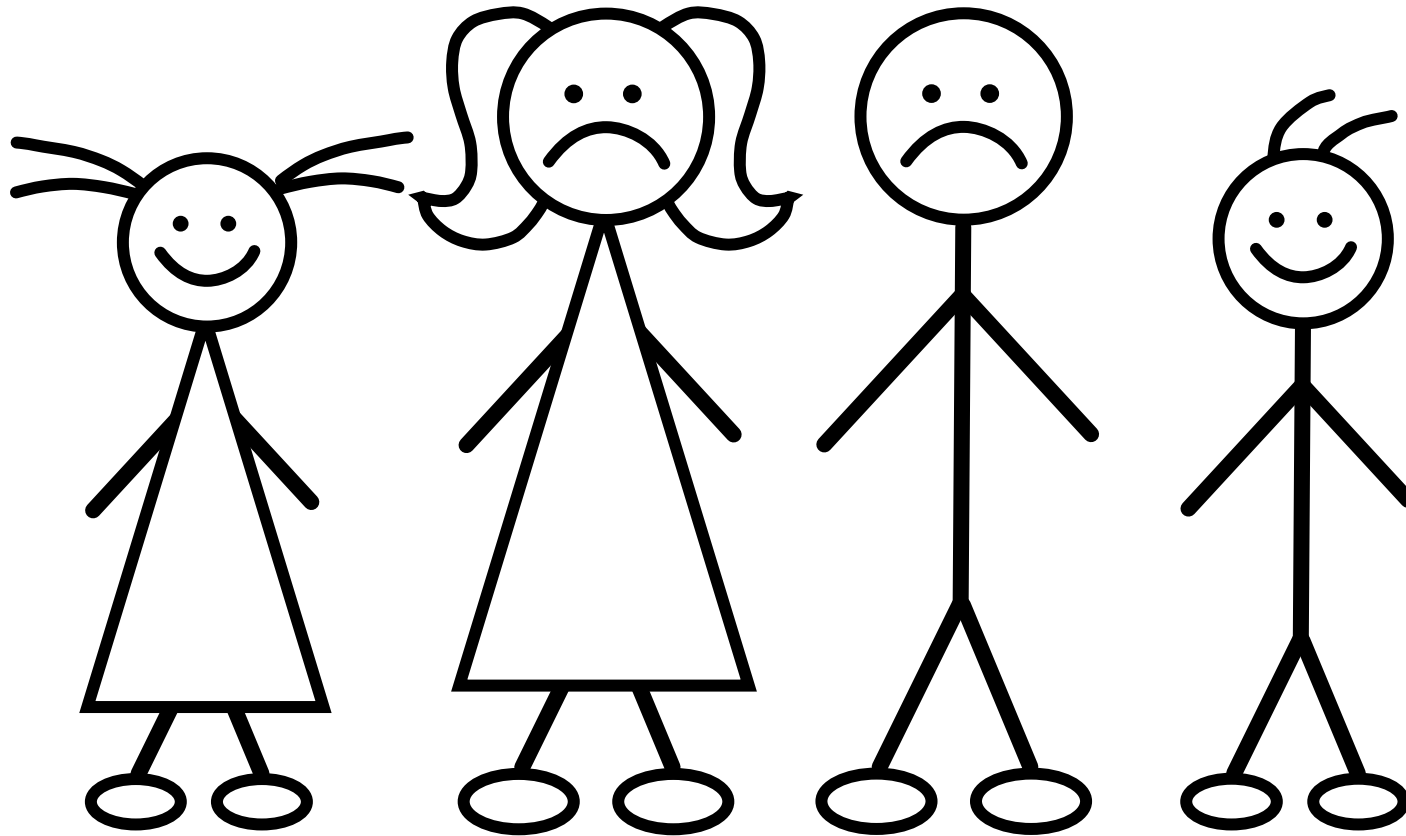




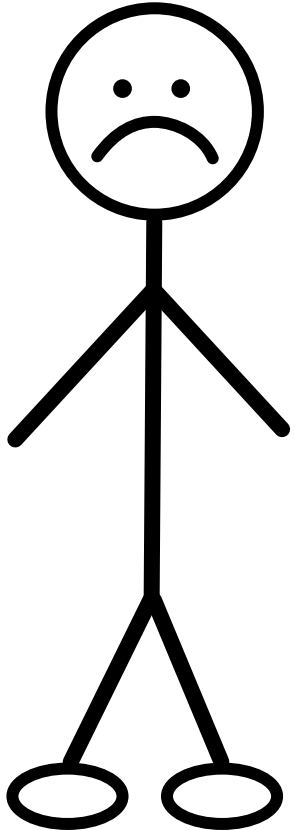
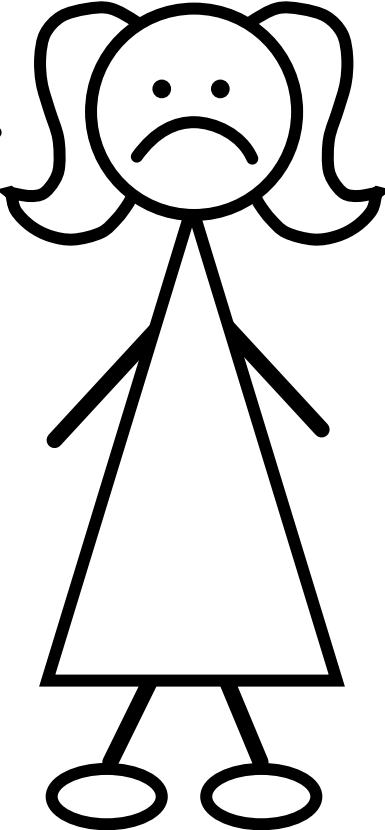
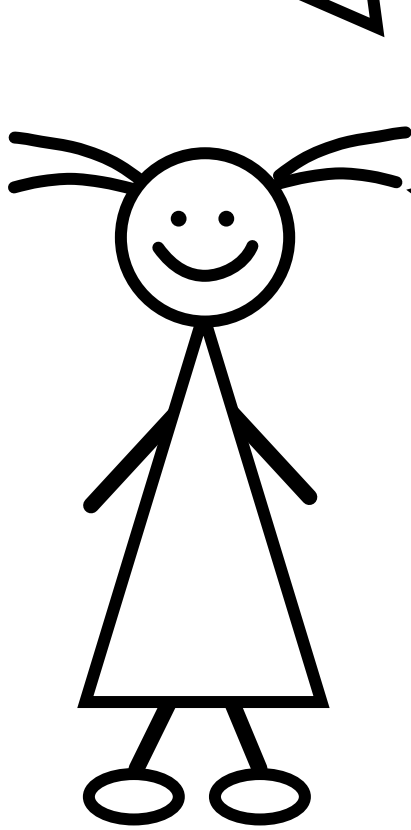
Computer
Scientist



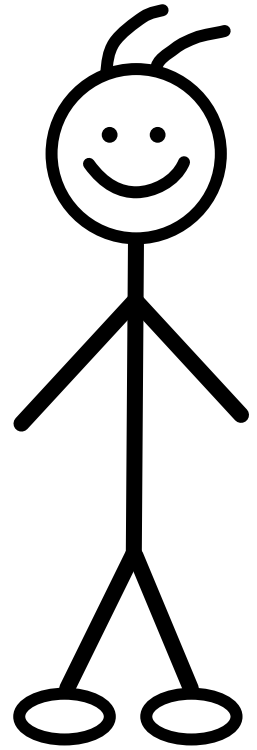
Data
Scientist

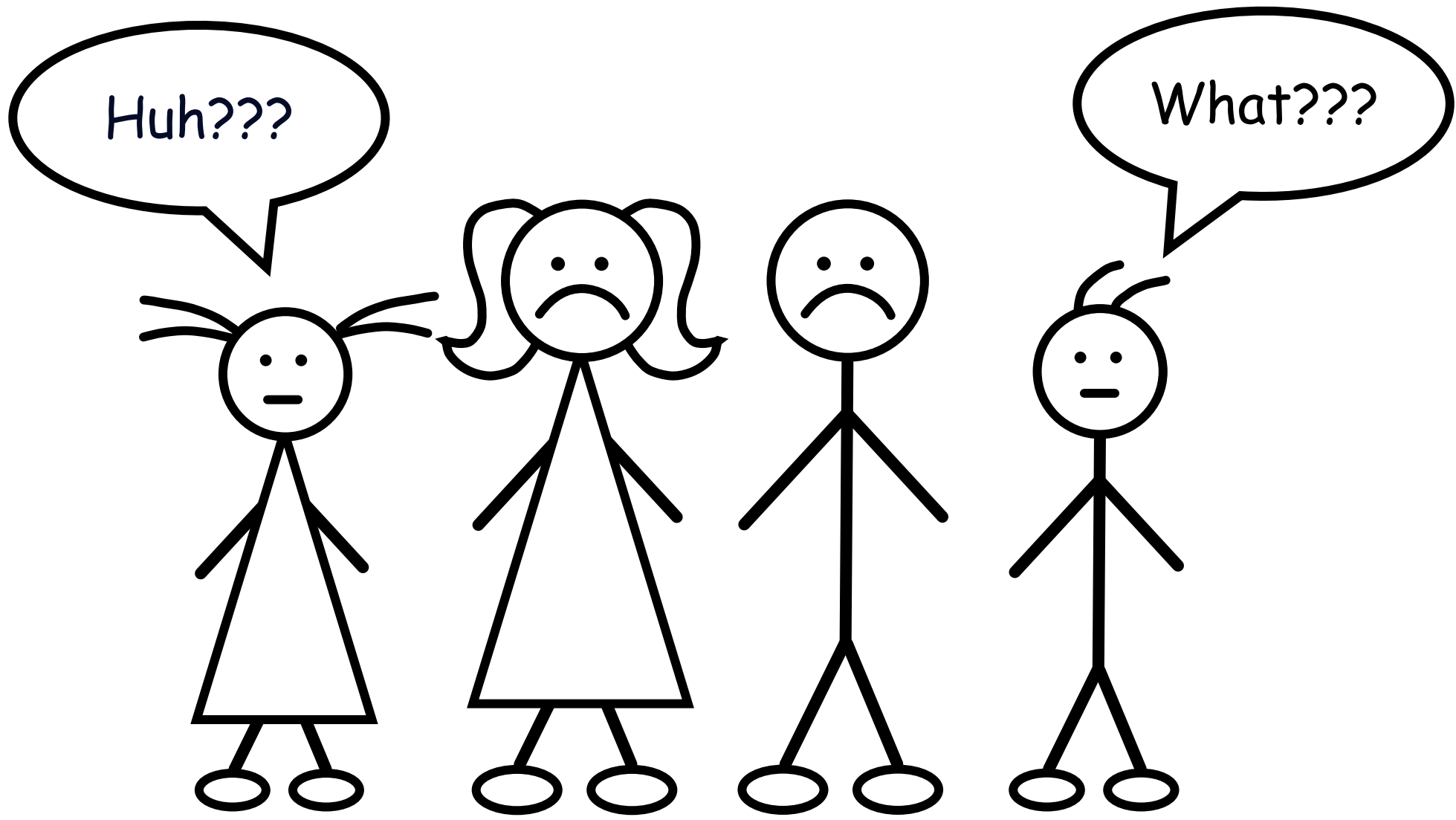


No jobs :(



Nope, none at all :(

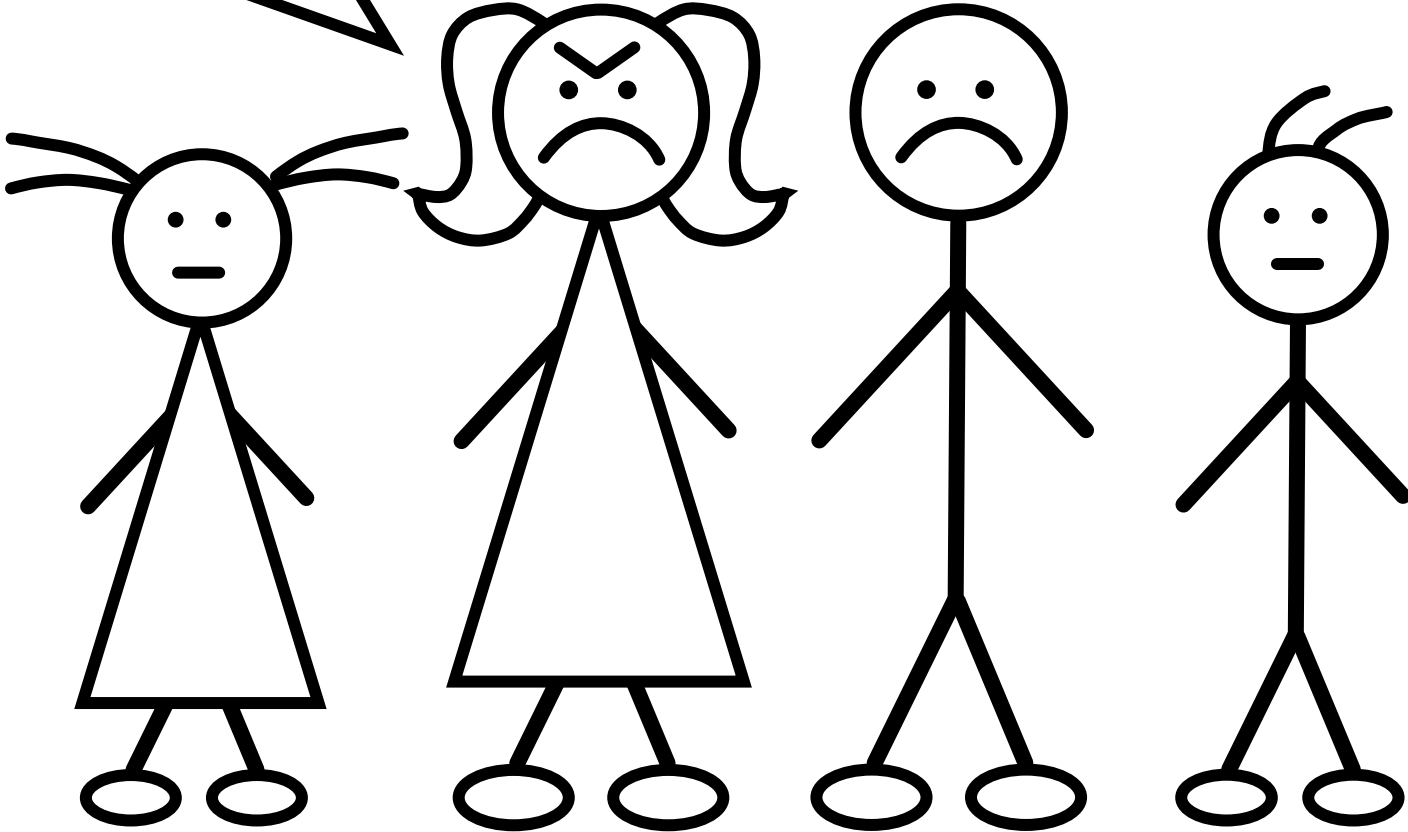


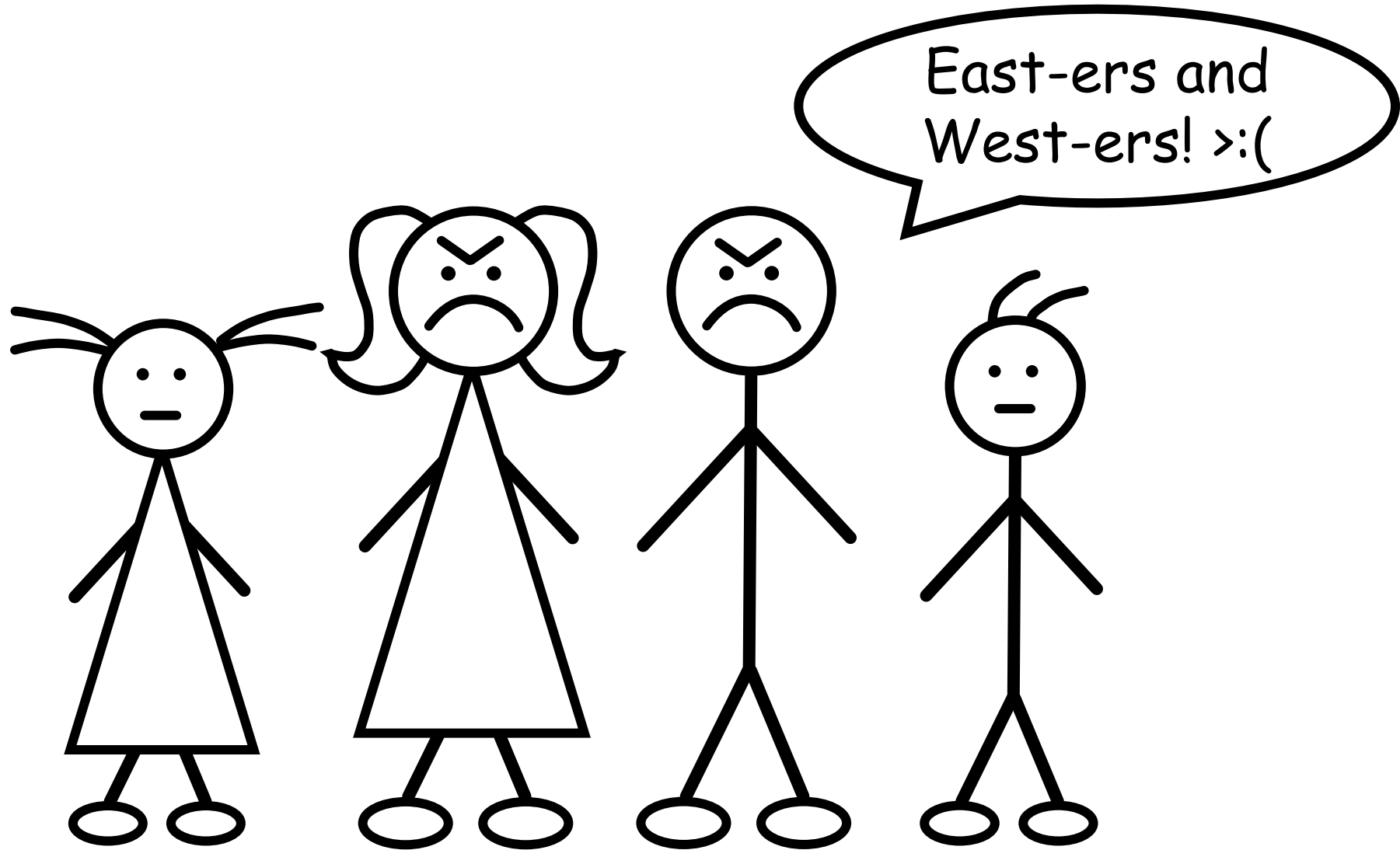


Huh???

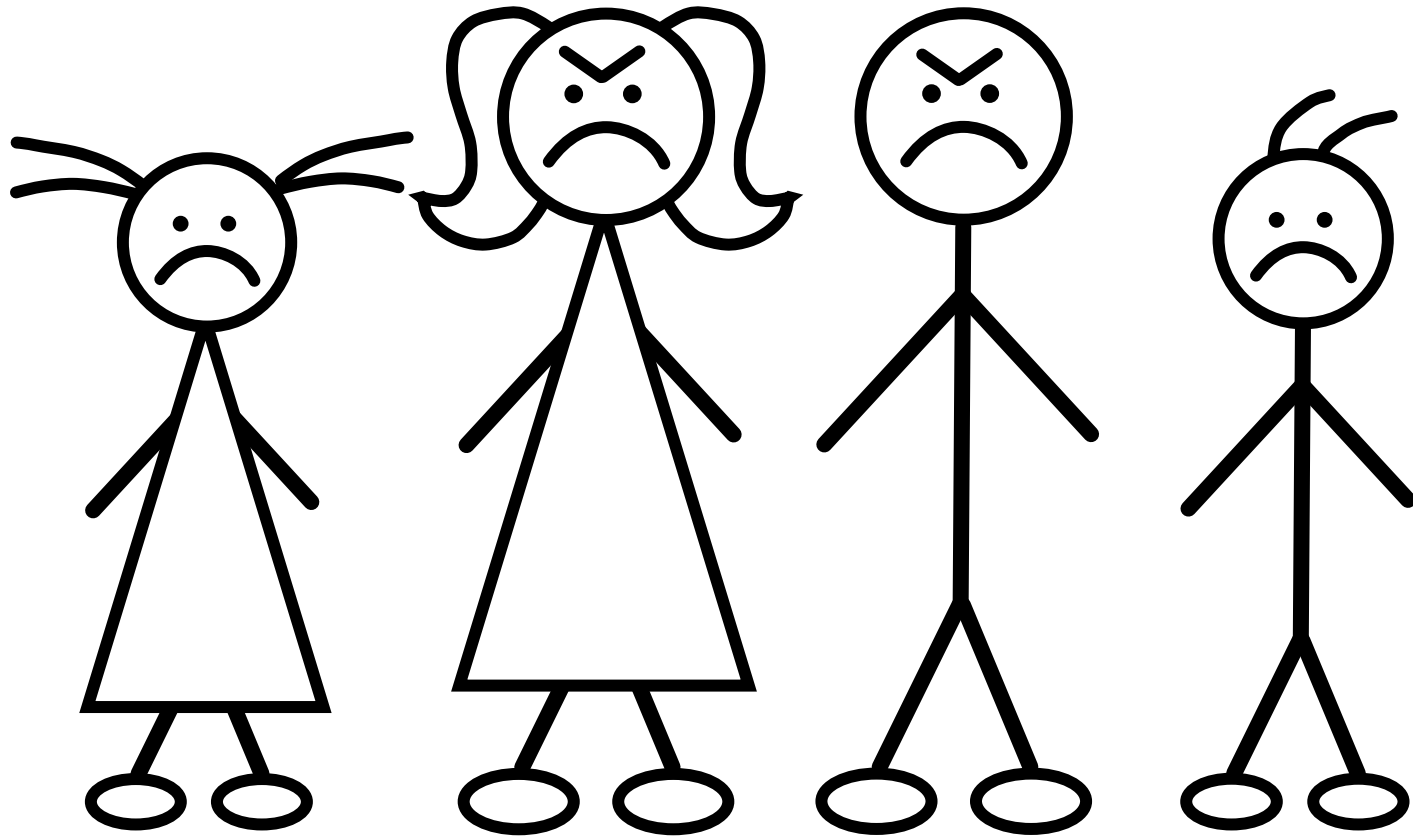
What???

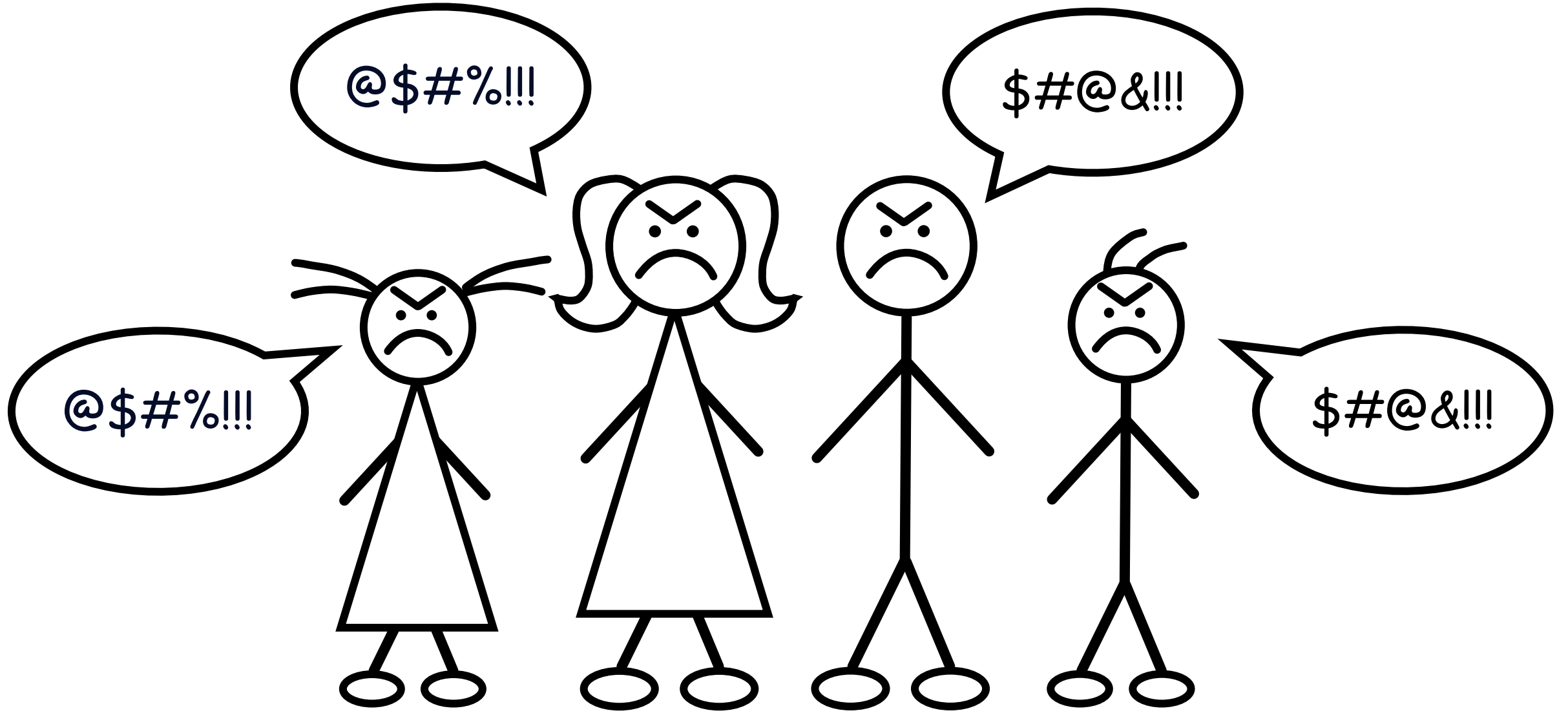
Alpha-Ilenials
blah, blah >:(

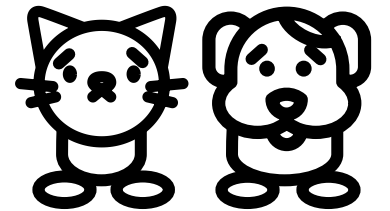
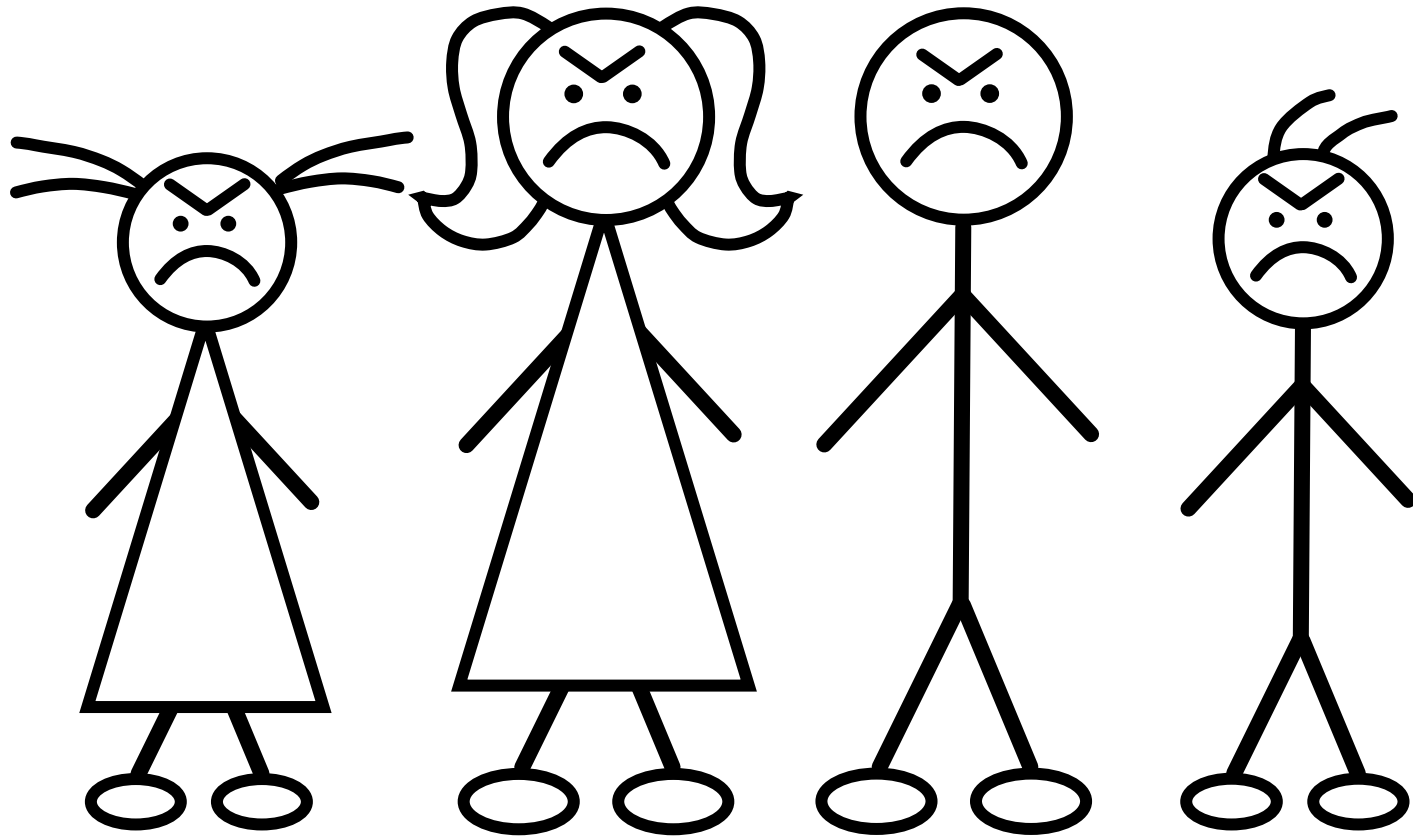




East-ers and
West-ers! >:(







Benefits



Automated labor





Autonomous vehicles



A quadcopter drone is shown in flight against a clear blue sky. The drone is carrying a brown cardboard box suspended from its underside. In the background, a city skyline is visible, featuring several buildings, including a prominent blue glass skyscraper. The drone's propellers are blurred, indicating it is in motion.

Delivery drones



Automated decisions

A worker wearing a yellow hard hat and safety glasses is looking at a virtual 3D model of a mechanical assembly. The model is semi-transparent, showing internal components like pistons and valves. The worker is holding a tool, possibly a wrench, near the model. The background is a blurred industrial setting. A green banner is overlaid across the middle of the image, containing the text "Virtual and augmented reality". At the bottom of the image, there is a horizontal bar with several icons: a refresh symbol, a gear, a wrench, a book, a crossed wrench and screwdriver, an envelope, a telephone, and a home symbol.

Virtual and augmented reality

A white and grey dual-arm robotic hand is positioned over a kitchen stove. The robot's arms are extended, with the right hand hovering over a stainless steel pot with a lid on the right burner, and the left hand hovering over a black frying pan on the left burner. The stove is set on a dark countertop in front of a window with a view of greenery. A green banner with white text is overlaid across the middle of the image.

Domestic robots



Smart buildings





Smart cities



A diverse group of office workers in a modern workspace, focused on their laptops and collaborative work. The scene is brightly lit, likely from a large window on the left. In the foreground, a woman in a light blue shirt is seated at a white desk, working on a laptop. To her right, another person is seated at a desk with a large monitor and keyboard. In the background, a man and a woman are standing and looking at a laptop screen, while another man is seated at a desk in front of them. The overall atmosphere is professional and collaborative.

High-tech jobs

Problems



Unemployment

A group of five people in business attire sitting in a row on chairs, representing job seekers. From left to right: a man in a dark suit, a man in a tan suit, a woman in a red suit holding a white paper, a woman in a grey checkered blazer holding a smartphone, and a man in a dark suit holding a tablet. The background is a plain, light-colored wall.

Simple, repetitive,
and costly jobs

Unemployment

A group of five business professionals (three men and two women) are sitting in a row on chairs. They are dressed in professional attire. The image is overlaid with two orange text boxes at the top and a larger orange text box at the bottom. The background is a plain, light-colored wall and floor.

Simple, repetitive,
and costly jobs

600 million jobs
lost by 2030

Unemployment



Simple, repetitive,
and costly jobs

600 million jobs
lost by 2030

700 million new
jobs by 2030

Unemployment



Simple, repetitive,
and costly jobs

600 million jobs
lost by 2030

700 million new
jobs by 2030

Unemployment

Labor market is
unprepared



Simple, repetitive,
and costly jobs

600 million jobs
lost by 2030

700 million new
jobs by 2030

Unemployment

Labor market is
unprepared

Society is
unprepared



Simple, repetitive,
and costly jobs

600 million jobs
lost by 2030

700 million new
jobs by 2030

Unemployment

Labor market is
unprepared

Society is
unprepared

Many people
left behind

A woman with long dark hair, wearing a light grey sweater, is sitting on a grey sofa. She is looking at a laptop screen with a serious and slightly concerned expression. Her hand is resting on the laptop keyboard. The background is a blurred indoor setting with a plant and a shelf.

Synthetic media

Indistinguishable
from humans

Synthetic media





Indistinguishable
from humans

Propaganda and
misinformation

Synthetic media

A woman with dark hair, wearing a grey sweater, is sitting on a grey couch. She is looking down at a laptop screen. The background is slightly blurred, showing a window and some greenery.

Indistinguishable
from humans

Propaganda and
misinformation

Deep fake
technology

Synthetic media



Indistinguishable
from humans

Propaganda and
misinformation

Deep fake
technology

Synthetic media

Deep nude
technology

Indistinguishable
from humans

Propaganda and
misinformation

Deep fake
technology

Synthetic media

Deep nude
technology

Synthetic
hyper-reality

Indistinguishable
from humans

Propaganda and
misinformation

Deep fake
technology

Synthetic media

Deep nude
technology

Synthetic
hyper-reality

Unhealthy virtual
relationships

Demographics

Age: 20
Gender: Female
Location: Chicago
Education: MS in Edu.
Occupation: Teacher
Income: \$35,100

Online Behavior

Last Visit: 2/15/2020
Last Purchase: 1/1/2020
Last Purchase: Shoes

In-app Behavior

Usage: 15 sessions / mo

Lack of agency and autonomy

Personality: Introverted
Lifestyle: Active
Interests: Dancing
Opinions: Liberal
Beliefs: Christian

Offline Behavior

Last In-store: 8/2/2019
Last purchase: Socks

Social Media

Last mention: 8/2/2019
Sentiment: Positive

Likes = behavioral insights

D
A
G
L
E
Occupation: Teacher
Income: \$35,100

P
P
Lifestyle: Active
Interests: Dancing
Opinions: Liberal
Beliefs: Christian

Lack of agency and autonomy

Online Behavior

Last Visit: 2/15/2020
Last Purchase: 1/1/2020
Last Purchase: Shoes

In-app Behavior

Usage: 15 sessions / mo

Offline Behavior

Last In-store: 8/2/2019
Last purchase: Socks

Social Media

Last mention: 8/2/2019
Sentiment: Positive

Likes = behavioral insights

Manipulate our decisions

Online Behavior
Last Visit: 2/15/2020
Last Purchase: 1/1/2020
Last Purchase: Shoes

In-app Behavior
Usage: 15 sessions / mo

Lack of agency and autonomy

Offline Behavior
Last In-store: 8/2/2019
Last purchase: Socks

Social Media
Last mention: 8/2/2019
Sentiment: Positive

D
A
G
L
E
O
Occupation: Teacher
Income: \$35,100

P
P
Lifestyle: Active
Interests: Dancing
Opinions: Liberal
Beliefs: Christian

Likes = behavioral insights

Manipulate our decisions

Target our weaknesses

Lack of agency and autonomy

D
A
G
L
E
Occupation: Teacher
Income: \$35,100

P
P
Lifestyle: Active
Interests: Dancing
Opinions: Liberal
Beliefs: Christian

In-app Behavior
Usage: 15 sessions / mo

Offline Behavior
Last In-store: 8/2/2019
Last purchase: Socks

Social Media
Last mention: 8/2/2019
Sentiment: Positive

Likes = behavioral insights

Manipulate our decisions

Target our weaknesses

Lack of agency and autonomy

Surveillance capitalism

D
A
G
L
E
Occupation: Teacher
Income: \$35,100

P
P
Lifestyle: Active
Interests: Dancing

O
B

In-app Behavior
Usage: 15 sessions / mo

Offline Behavior
Last In-store: 8/2/2019
Last purchase: Socks

Social Media
Last mention: 8/2/2019
Sentiment: Positive

Likes = behavioral insights

Manipulate our decisions

Target our weaknesses

Lack of agency and autonomy

Surveillance capitalism

Lack of freedom

D
A
G
L
E
Occupation: Teacher
Income: \$35,100

P
P
Lifestyle: Active
Interests: Dancing
O
B

In-app Behavior
Usage: 15 sessions / mo

Offline Behavior
Last In-store: 8/2/2019
Last purchase: Socks

Social Media
Last mention: 8/2/2019
Sentiment: Positive

Likes = behavioral insights

Manipulate our decisions

Target our weaknesses

Lack of agency and autonomy

Surveillance capitalism

Lack of freedom

Forced compliance

D
A
G
L
E
Occupation: Teacher
Income: \$35,100

P
P
Lifestyle: Active
Interests: Dancing

O
B

In-app Behavior
Usage: 15 sessions / mo

Offline Behavior
Last location: 9/2/2010

A 3D illustration featuring several stacks of US dollar bills. The stacks are arranged in a row, with the tallest stack on the right and the shortest on the left. Blue, stylized human figures are placed on top of the stacks, with the tallest figure on the tallest stack and the shortest on the shortest stack. The background is white, and the stacks cast soft shadows on the surface below. An orange horizontal bar is overlaid across the middle of the image, containing the text 'Economic impacts' in white.

Economic impacts



Social stratification



Safety issues



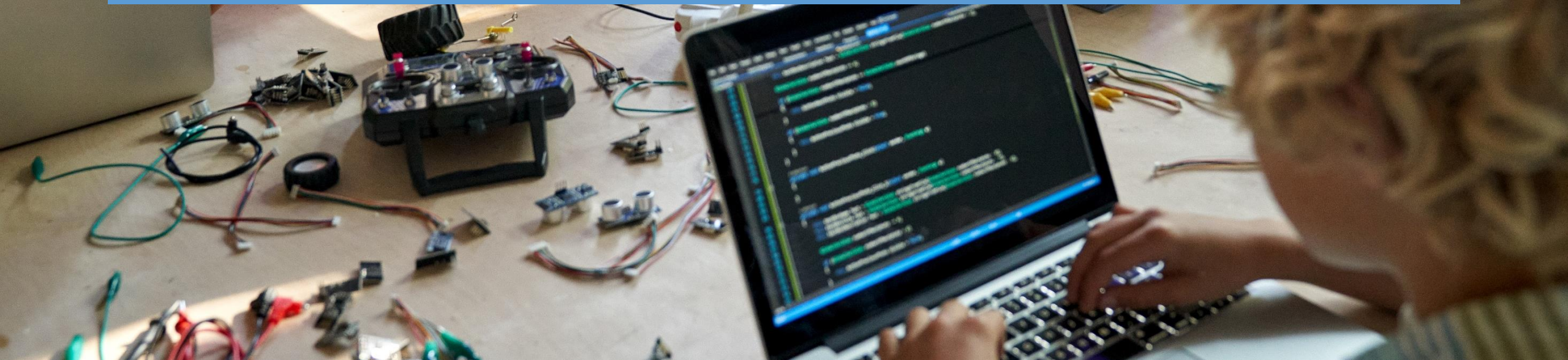
A silver autonomous weapon drone is shown in flight against a blue sky with light clouds. The drone is viewed from a low angle, showing its nose, cockpit area, and wings. The wings have the letters 'FL' and 'N' visible. The drone is flying over a desert landscape with green and brown terrain. A large orange banner is overlaid across the center of the image, containing the text 'Autonomous weapons' in white.

Autonomous weapons

Solutions



Educate everyone



A group of diverse children in a workshop setting. One girl is holding a small robot, while others look on with interest. The background shows shelves with various supplies and a window with a view of greenery.

What is AI?

A close-up of a child's hands typing on a laptop keyboard. The screen displays a code editor with lines of code in various colors. The desk is cluttered with electronic components, wires, and a small robot.

Educate everyone

A photograph of several children in a classroom or workshop. In the foreground, a young boy with dark hair is looking down at something in his hands. Behind him, a girl with curly hair is also looking down. To the left, another girl is partially visible. The background shows a window with a view of greenery outside and some storage bins on a shelf.

What is AI?

Basics of
ML, DL, RL

A photograph of a child with blonde hair sitting at a desk, typing on a laptop. The laptop screen displays a code editor with lines of code in various colors. On the desk in front of the laptop, there are various electronic components, including a breadboard, wires, and a small robot-like device with a camera lens.

Educate everyone

The background of the slide is a collage of images. The top half shows three children looking at a small robot on a table. The bottom half shows a child's hands typing on a laptop keyboard, with the screen displaying code. The table is cluttered with various electronic components like wires, breadboards, and small modules.

What is AI?

Basics of
ML, DL, RL

Pros and cons
of AI

Educate everyone



What is AI?

Basics of
ML, DL, RL

Pros and cons
of AI

Educate everyone



Capabilities and
limitations



What is AI?

Basics of
ML, DL, RL

Pros and cons
of AI

Educate everyone

Capabilities and
limitations

Impacts of AI

A background image showing several children in a classroom setting. They are gathered around a table, working on a small robot. One child is holding the robot, while others are looking on. The scene is brightly lit, suggesting a window in the background. The children are wearing casual clothing, and the overall atmosphere is one of collaborative learning and hands-on education.

What is AI?

Basics of
ML, DL, RL

Pros and cons
of AI

Educate everyone

Capabilities and
limitations

Impacts of AI

Ethical issues



Upgrade skills





Job retraining
programs

Upgrade skills





Job retraining
programs

Retirement
options

Upgrade skills



Job retraining
programs

Retirement
options

Social
safety nets

Upgrade skills



Job retraining
programs

Retirement
options

Social
safety nets

Upgrade skills

Mandatory
higher education

A background image of a factory or industrial setting. A worker wearing a hard hat and safety glasses is visible on the left, looking towards a robotic arm. The scene is filled with industrial machinery, control panels, and various gauges. The overall lighting is bright and industrial.

Job retraining
programs

Retirement
options

Social
safety nets

Upgrade skills

Mandatory
higher education

Redistribute wealth
and resources



Job retraining
programs

Retirement
options

Social
safety nets

Upgrade skills

Mandatory
higher education

Redistribute wealth
and resources

Universal basic
income

A person is shown from the chest up, wearing a light blue shirt, looking down at a laptop. The laptop keyboard is visible at the bottom. A large, semi-transparent blue horizontal bar is centered across the image, containing the text "Verify everything" in white. Below the bar, a green shield icon with a white checkmark is superimposed over the laptop keyboard. The background is filled with a complex pattern of white and yellow circuit lines and nodes, suggesting a digital or technological theme.

Verify everything

Deep-fake
detection

Verify everything



Deep-fake
detection

Digital alibi
services

Verify everything



Deep-fake
detection

Digital alibi
services

Information
verification

Verify everything



Deep-fake
detection

Digital alibi
services

Information
verification

Verify everything

Blockchain for
everything

Deep-fake
detection

Digital alibi
services

Information
verification

Verify everything

Blockchain for
everything

Default mode
of skepticism

Deep-fake
detection

Digital alibi
services

Information
verification

Verify everything

Blockchain for
everything

Default mode
of skepticism

Regulation of
synthetic media

A close-up photograph showing a person's hands holding a pair of heavily rusted, old-fashioned scissors. A thick, dark metal chain is wrapped around the blades of the scissors. To the left of the scissors, a black smartphone is lying on a white surface. The background is plain white. A blue horizontal bar is overlaid across the middle of the image, containing white text.

Maintain our agency and autonomy



Implement safety solutions



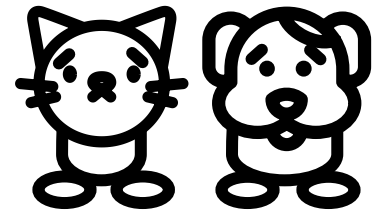
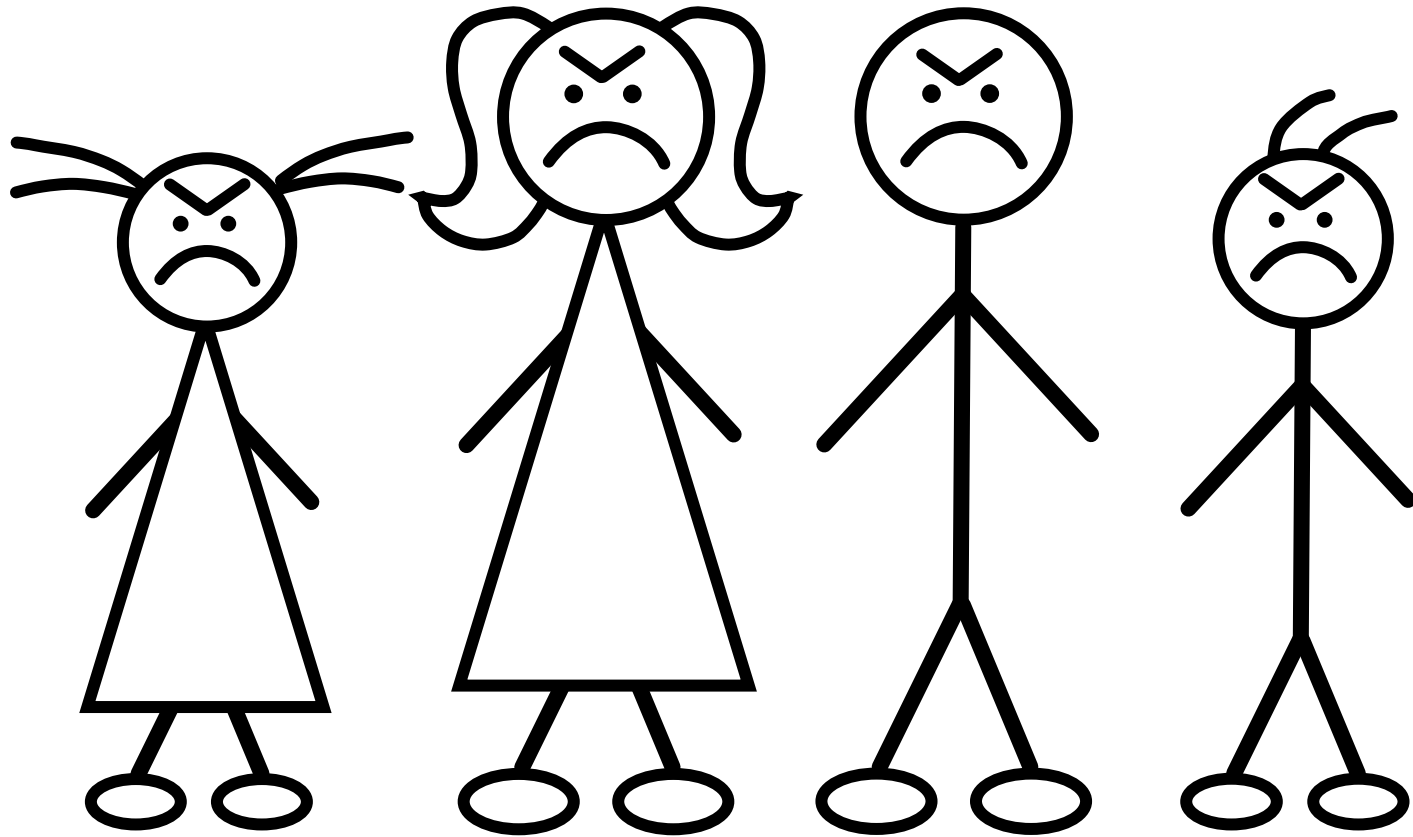


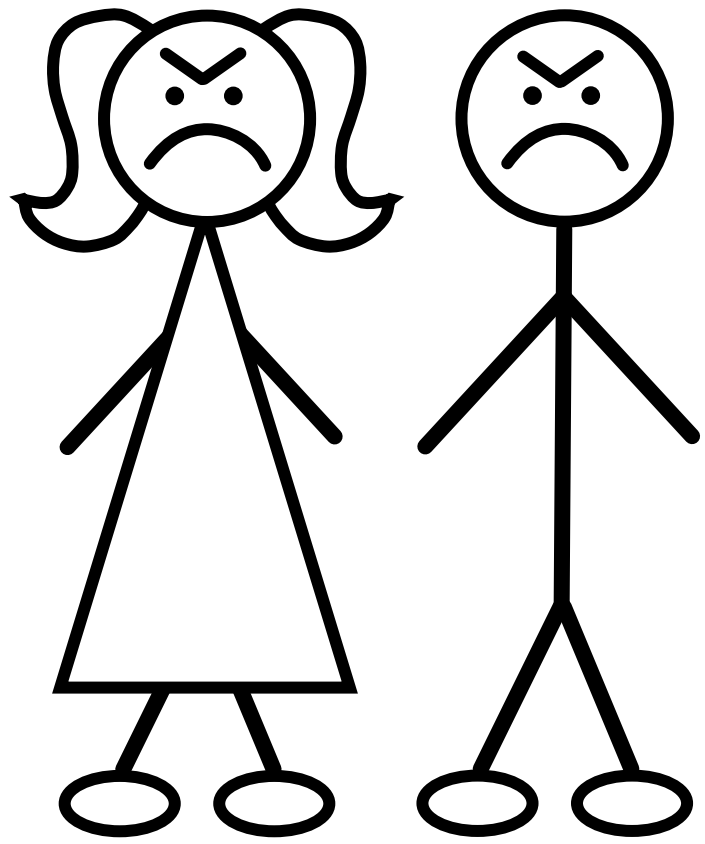
UNITED NATIONS  NATIONS UNIES

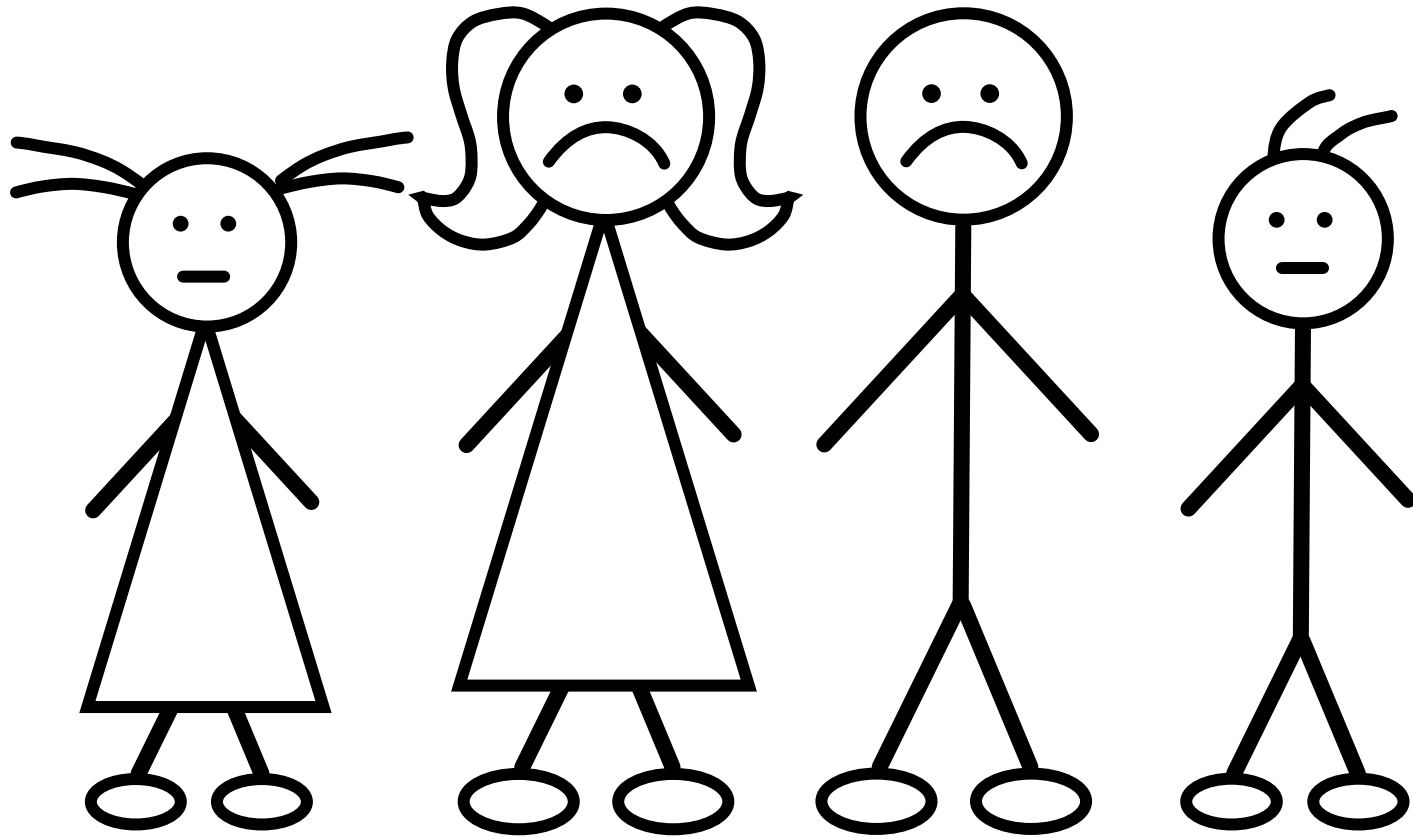
Ban autonomous weapons

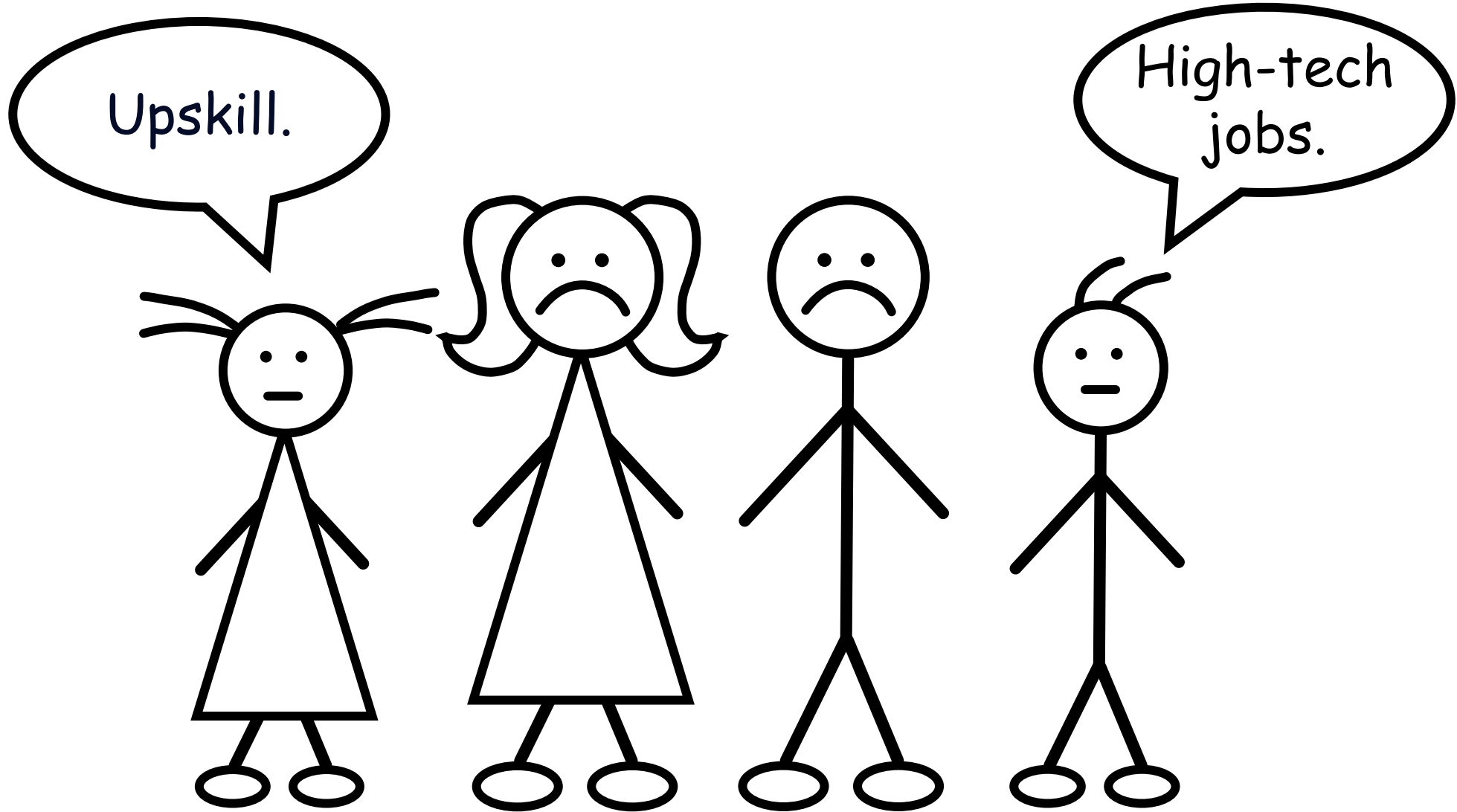
Alice and Bob in AI Land

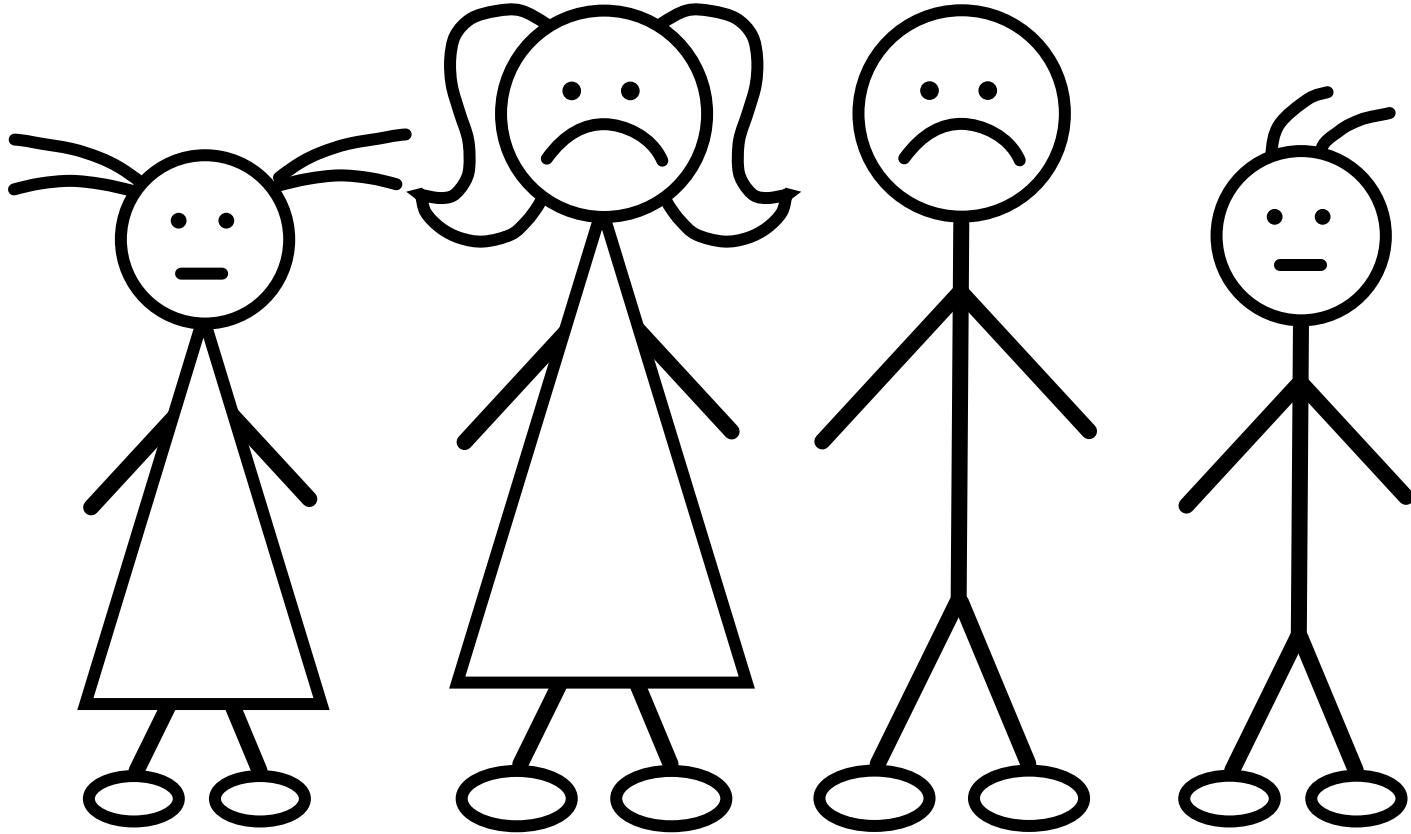
(continued ...)

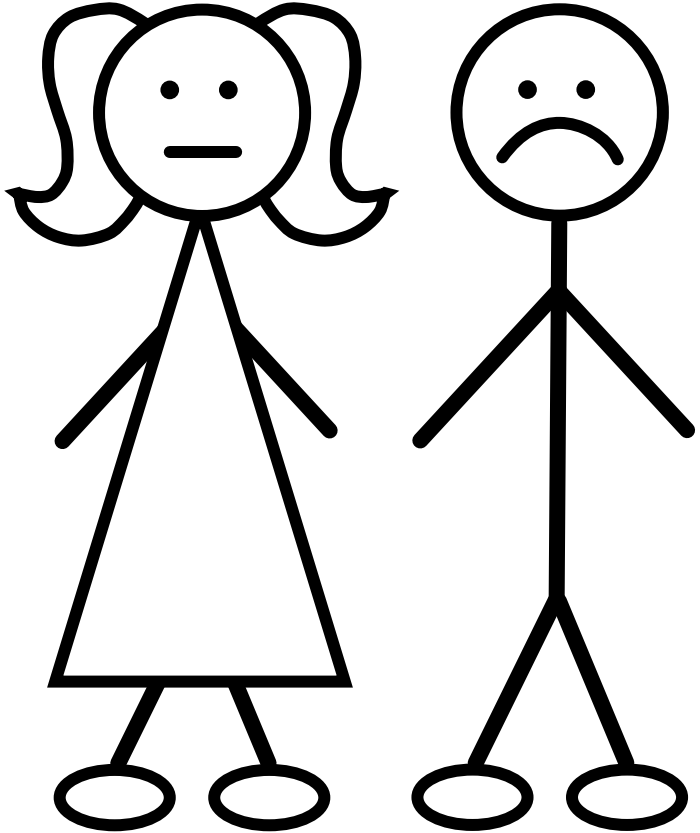


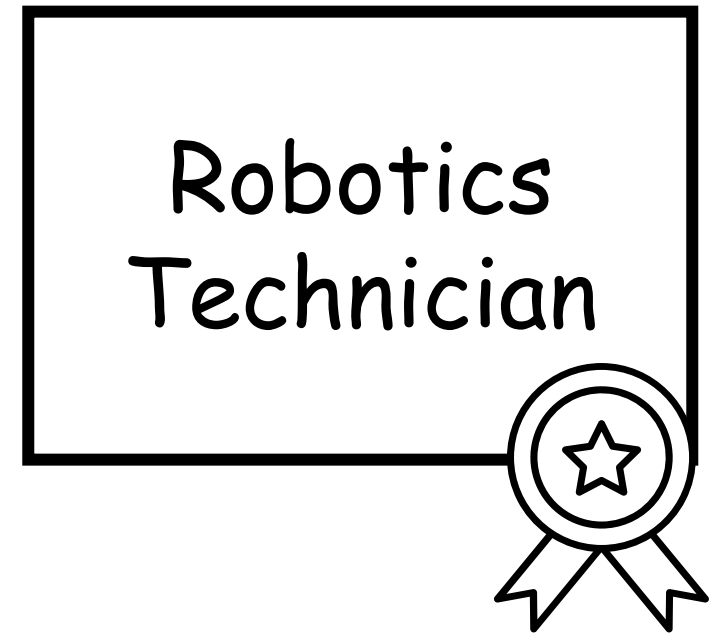
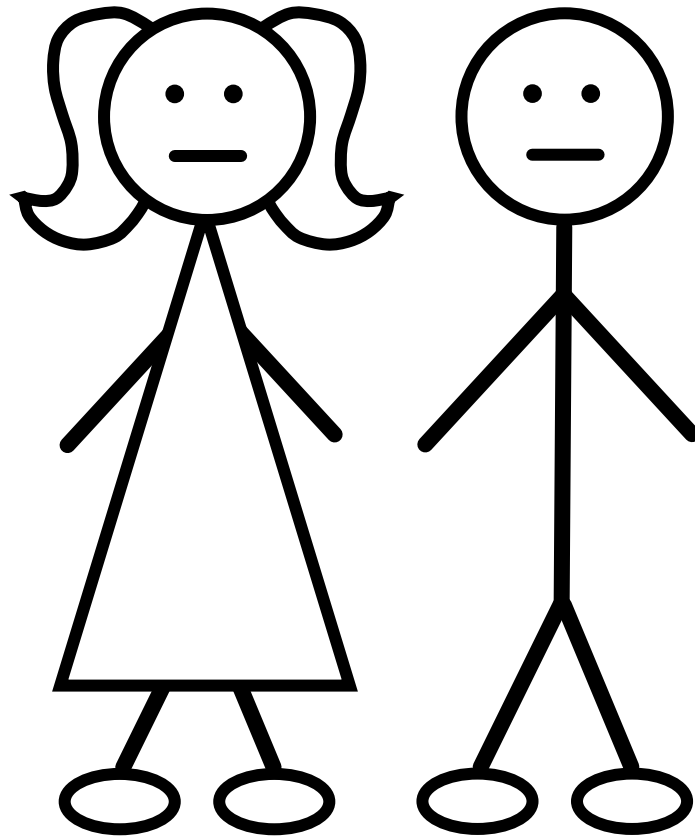


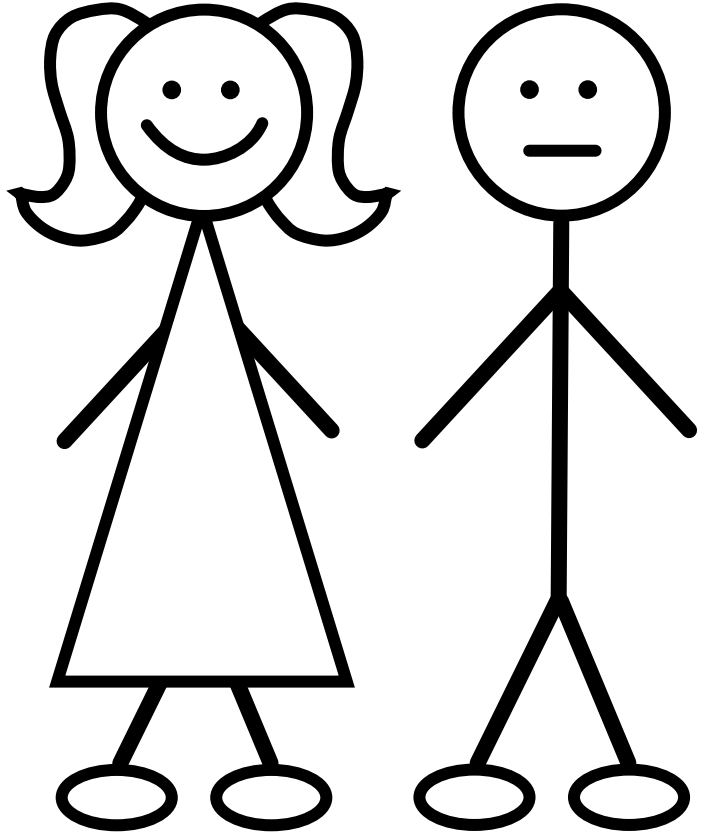


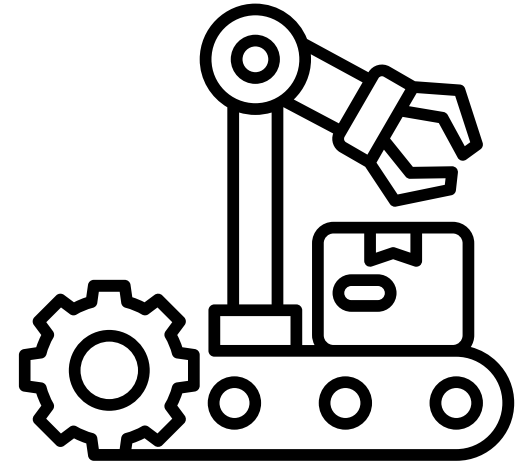
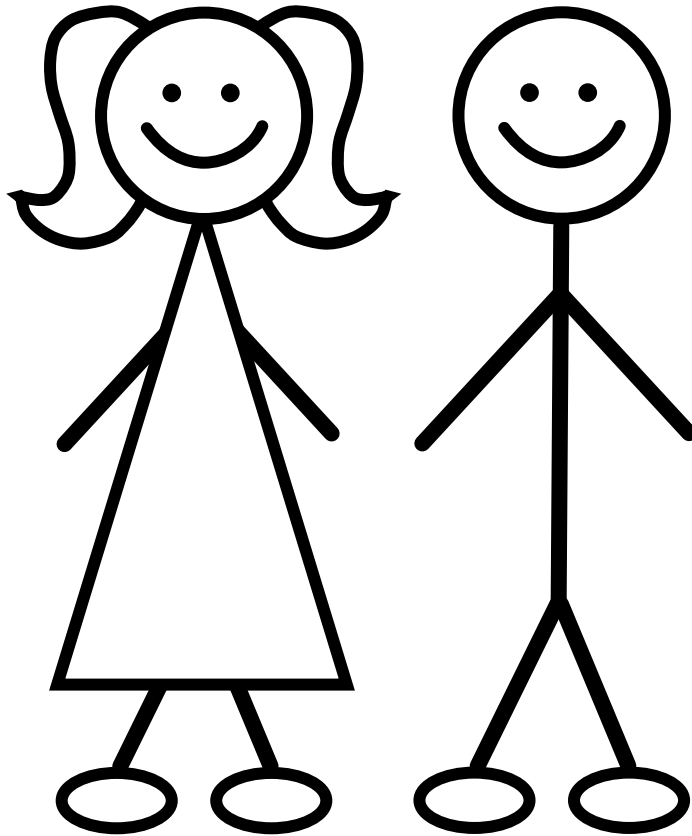


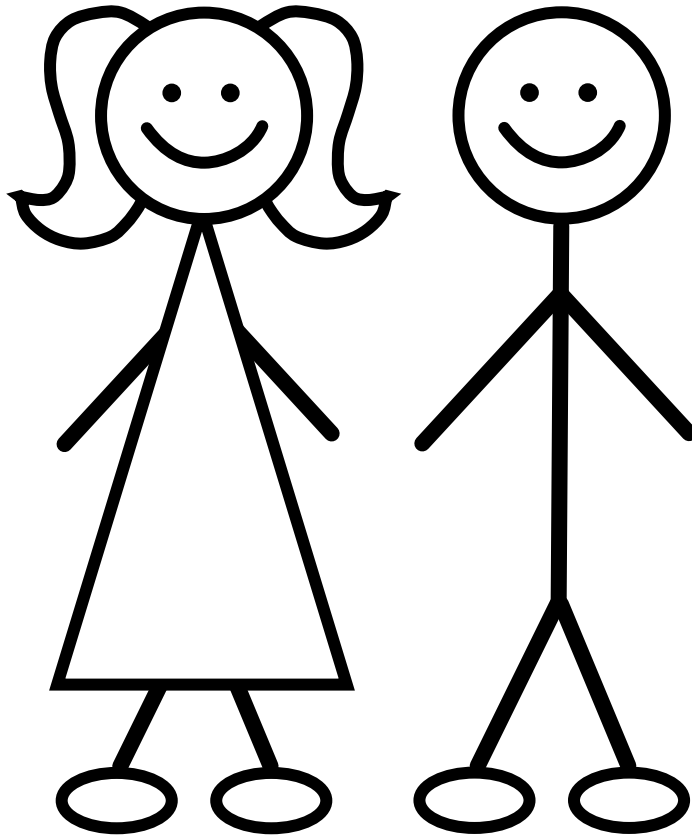


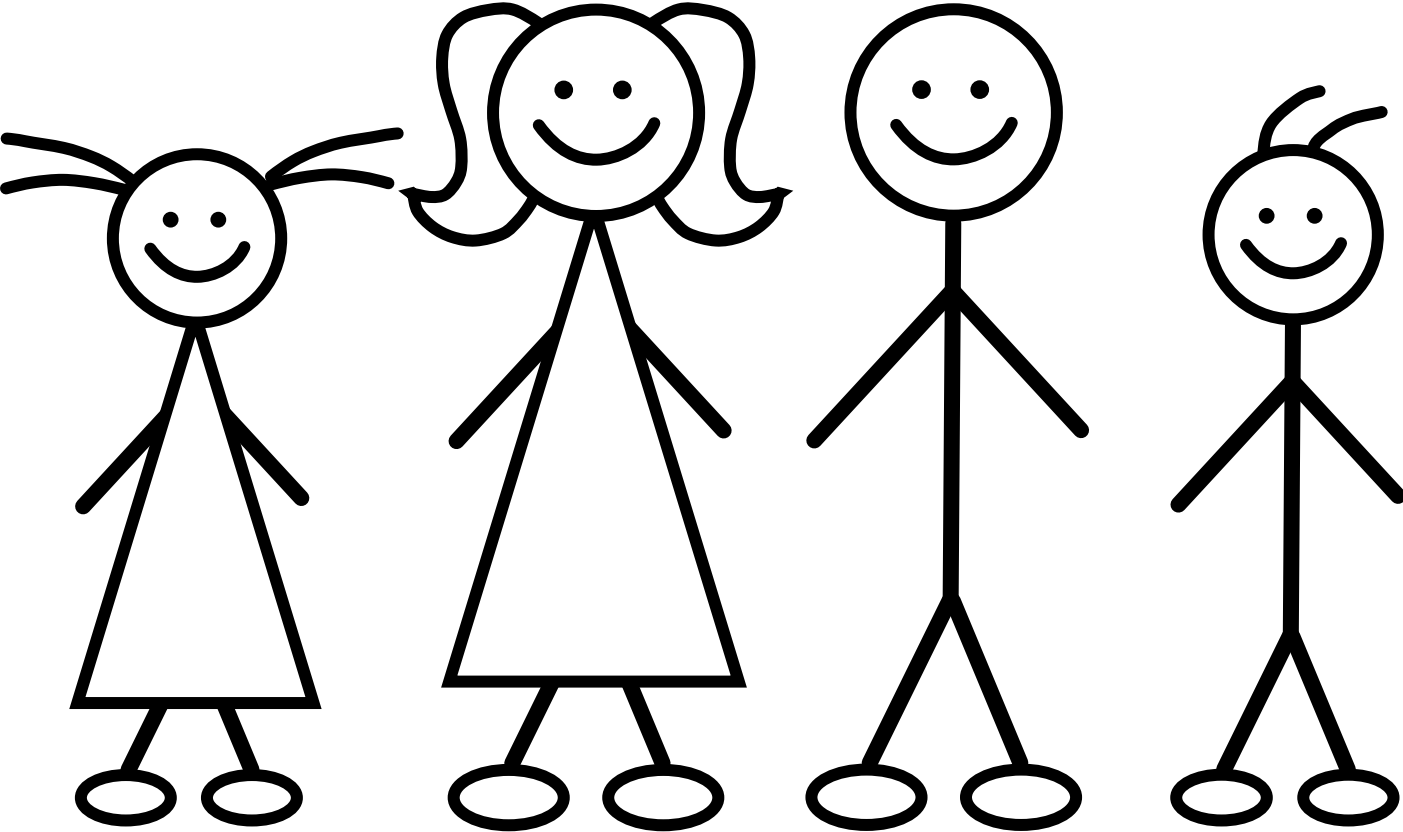




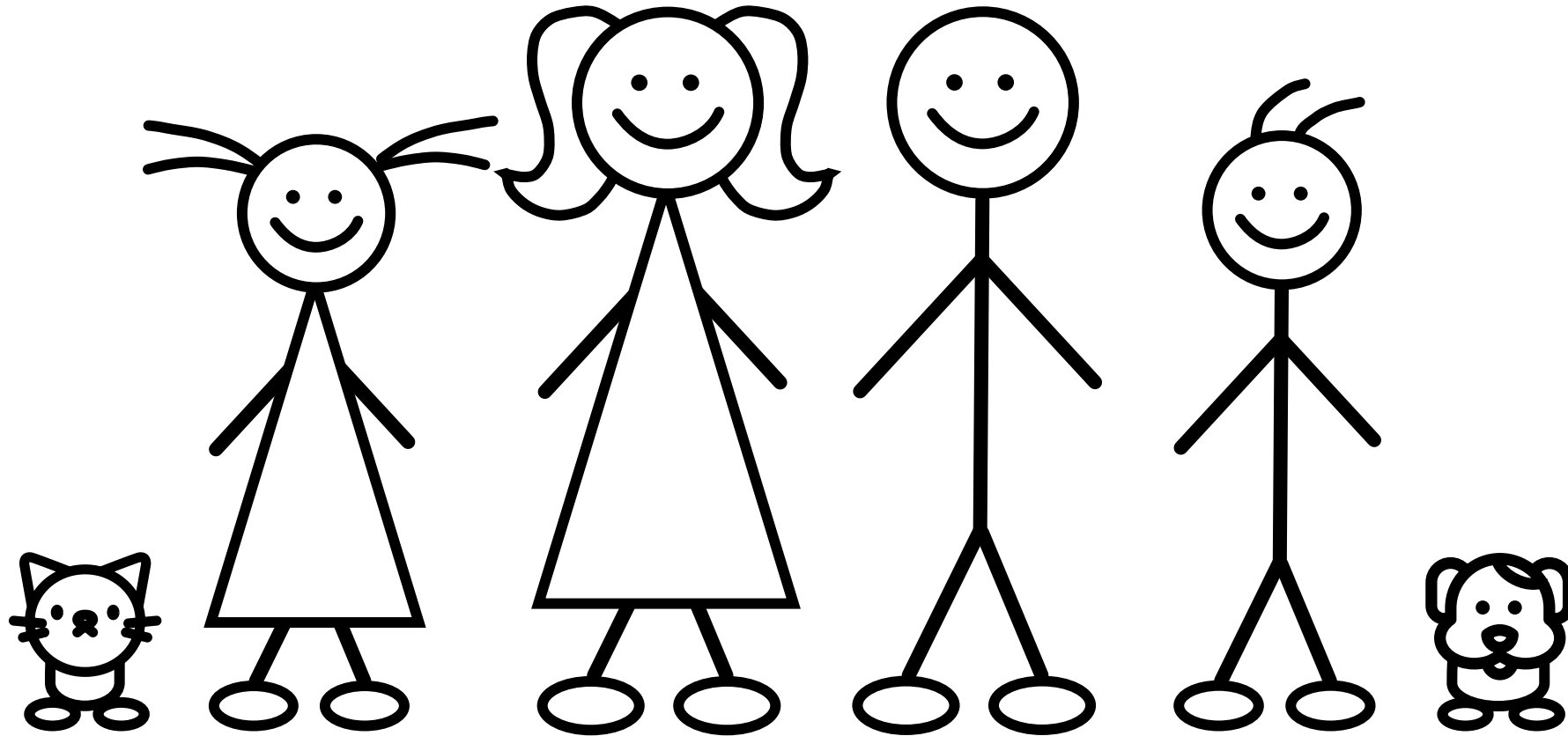








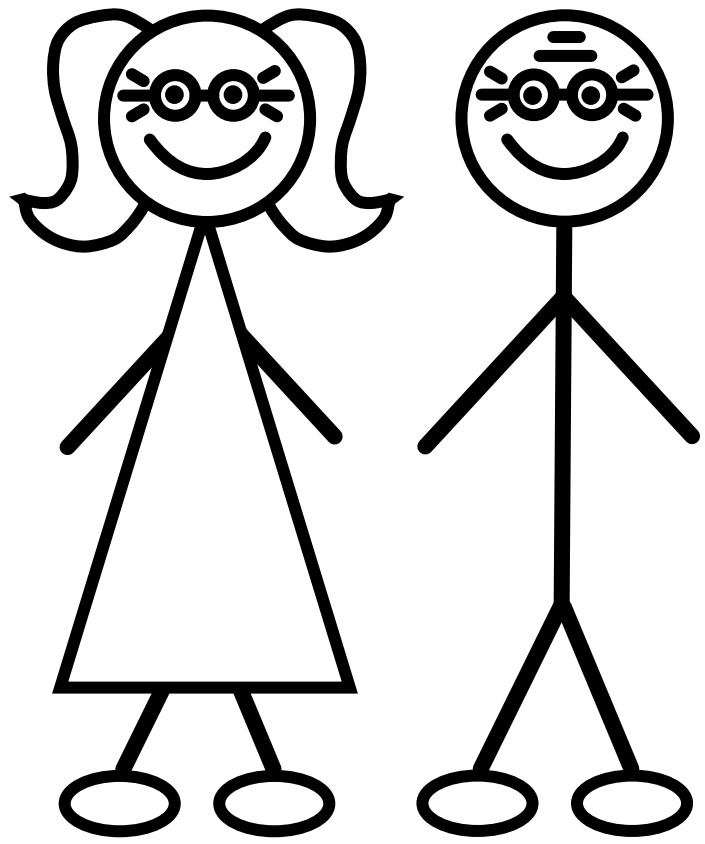
Happily ever after...

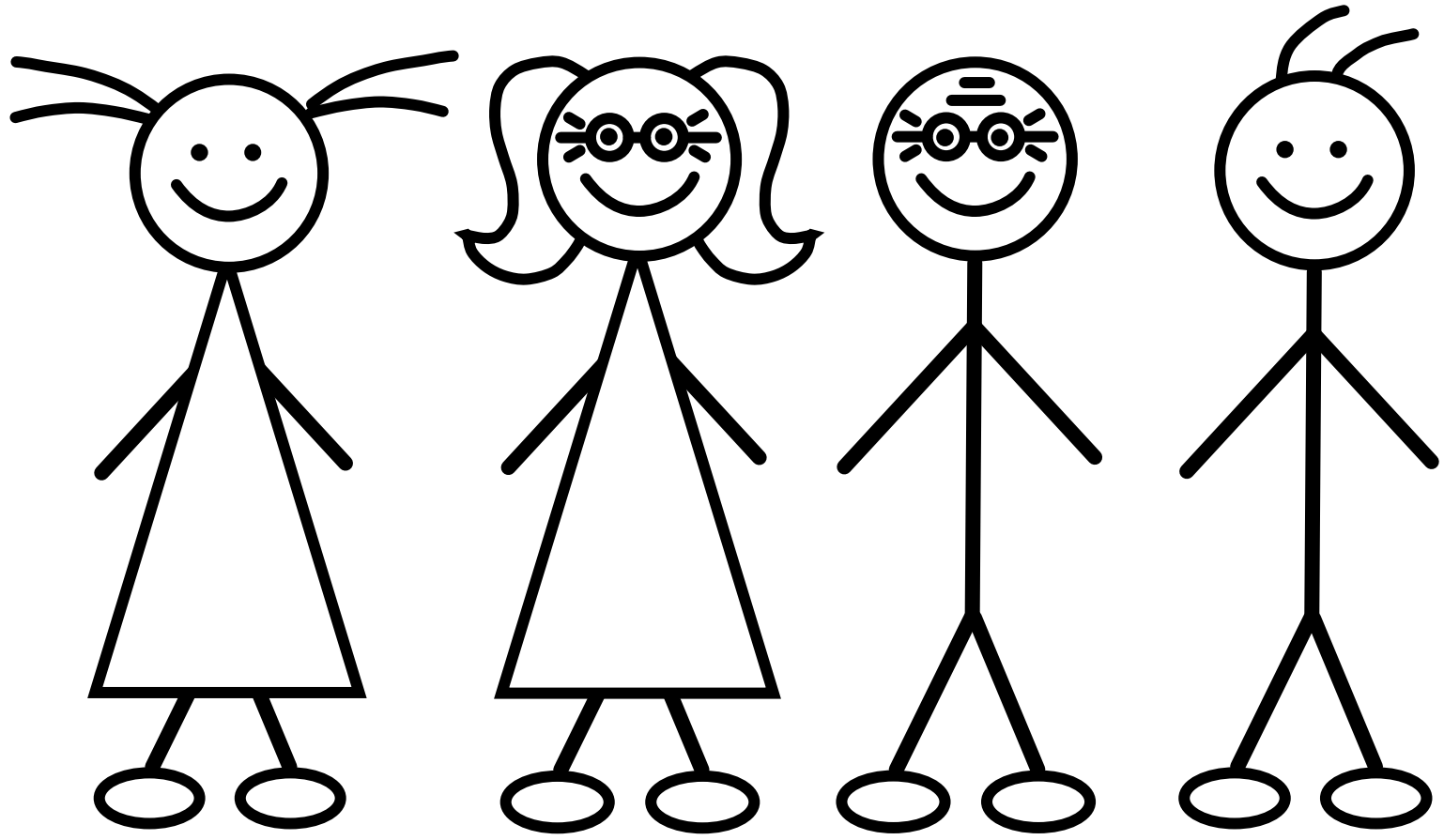


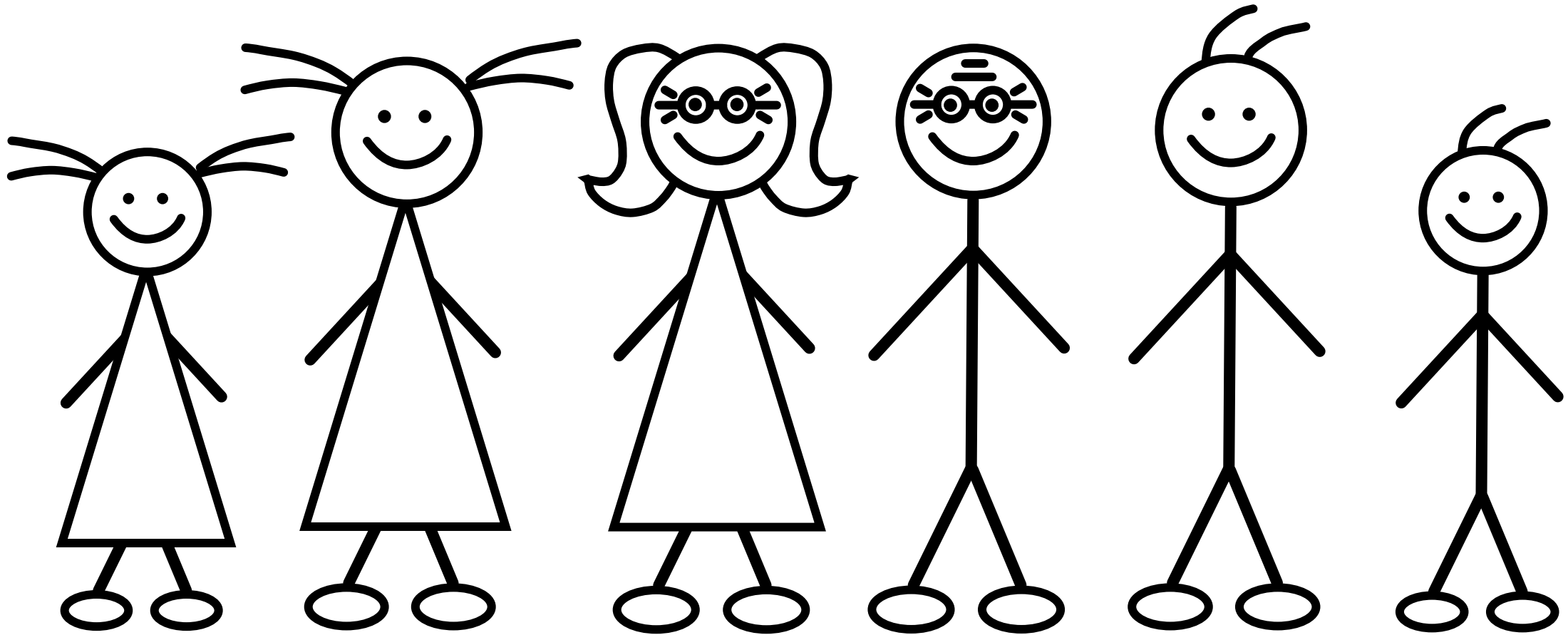
AI Ethics for The Future

Alice and Bob in AI Land

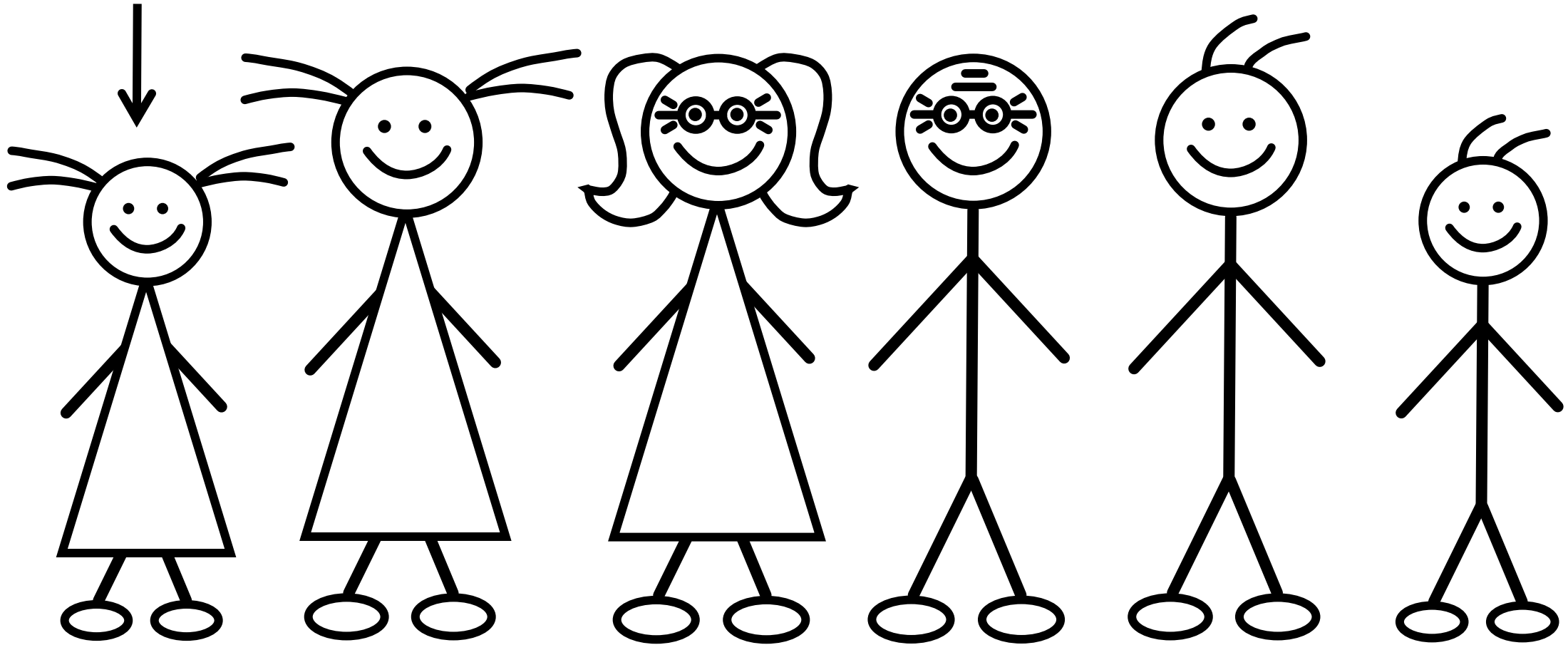
Chapter 3:
"AI Ethics of the Future"



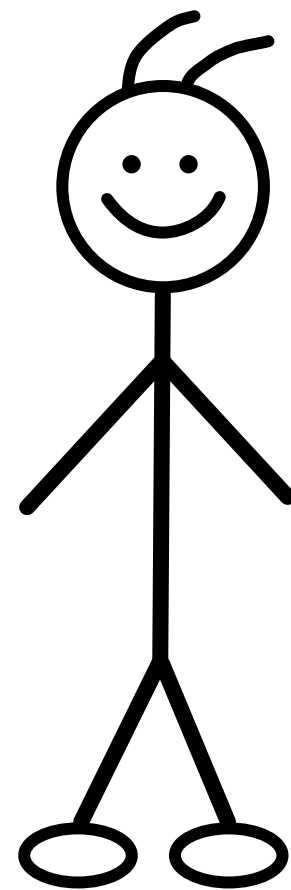
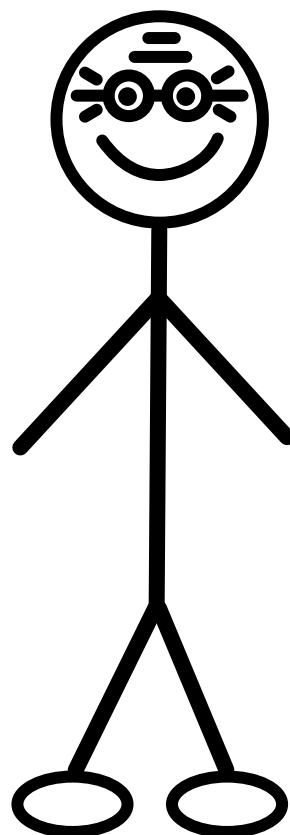
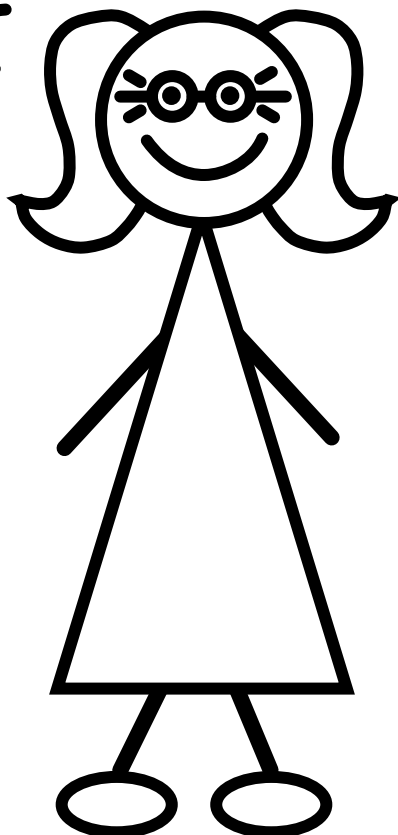
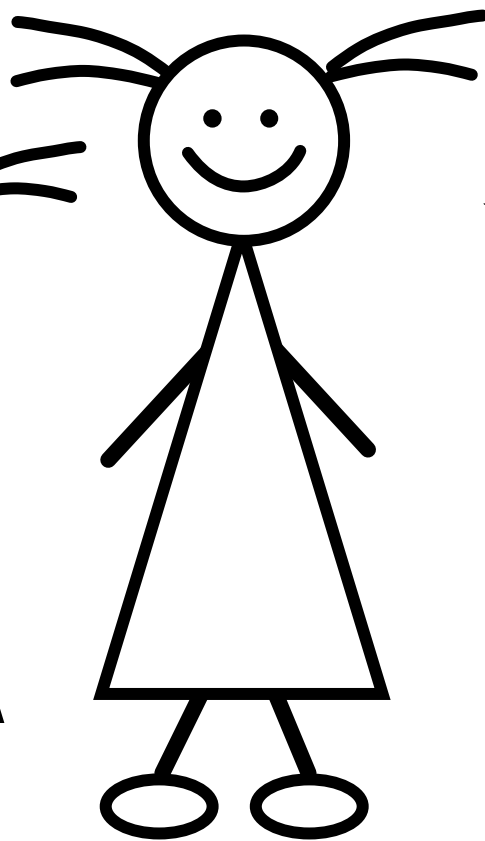
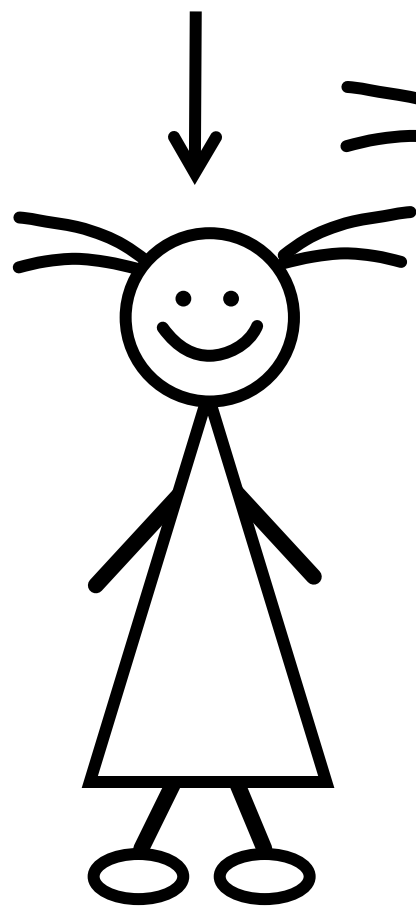




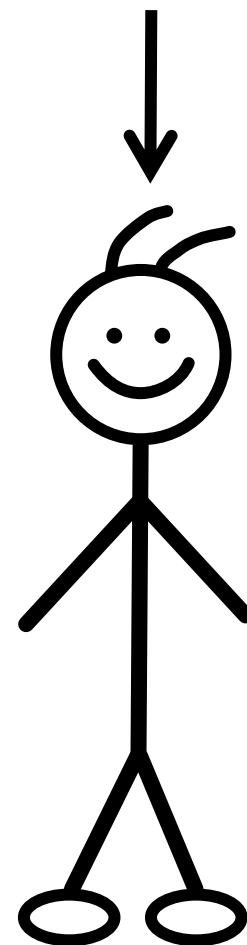
Emma

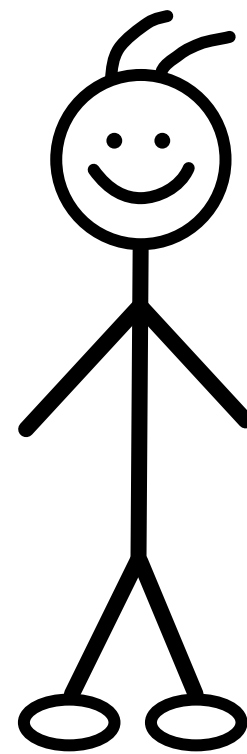
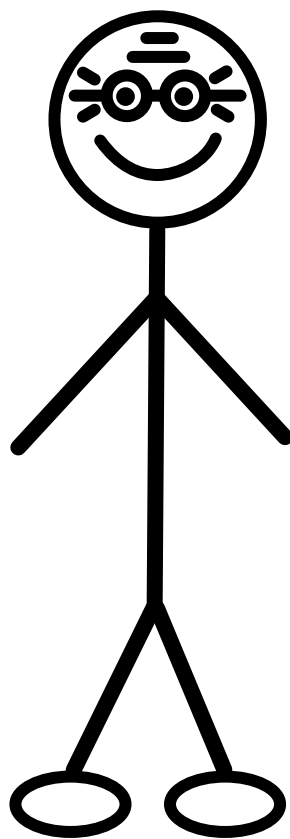
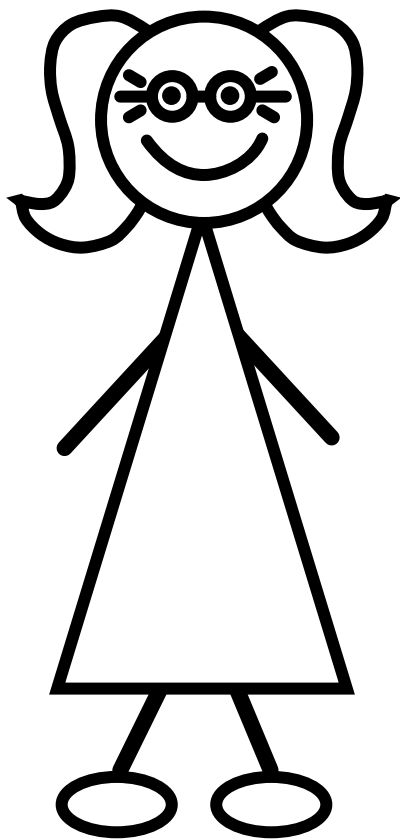
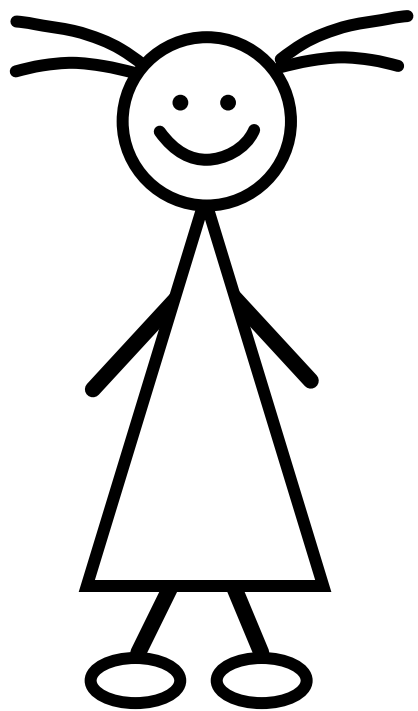


Emma

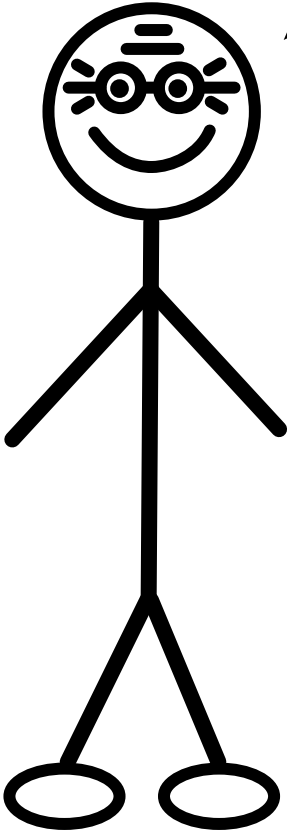
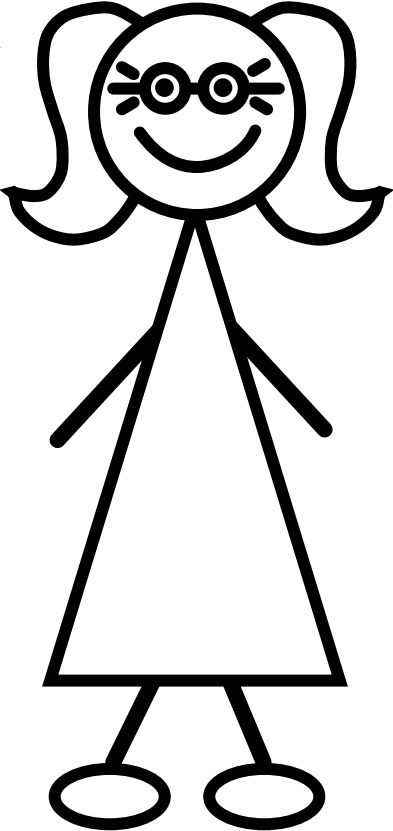
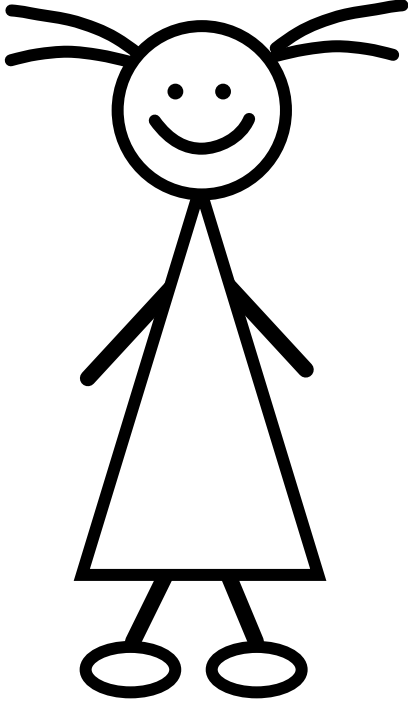


Fred

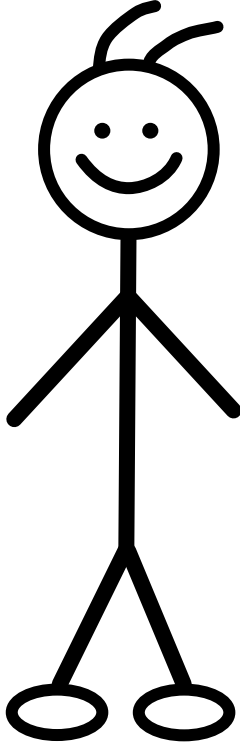




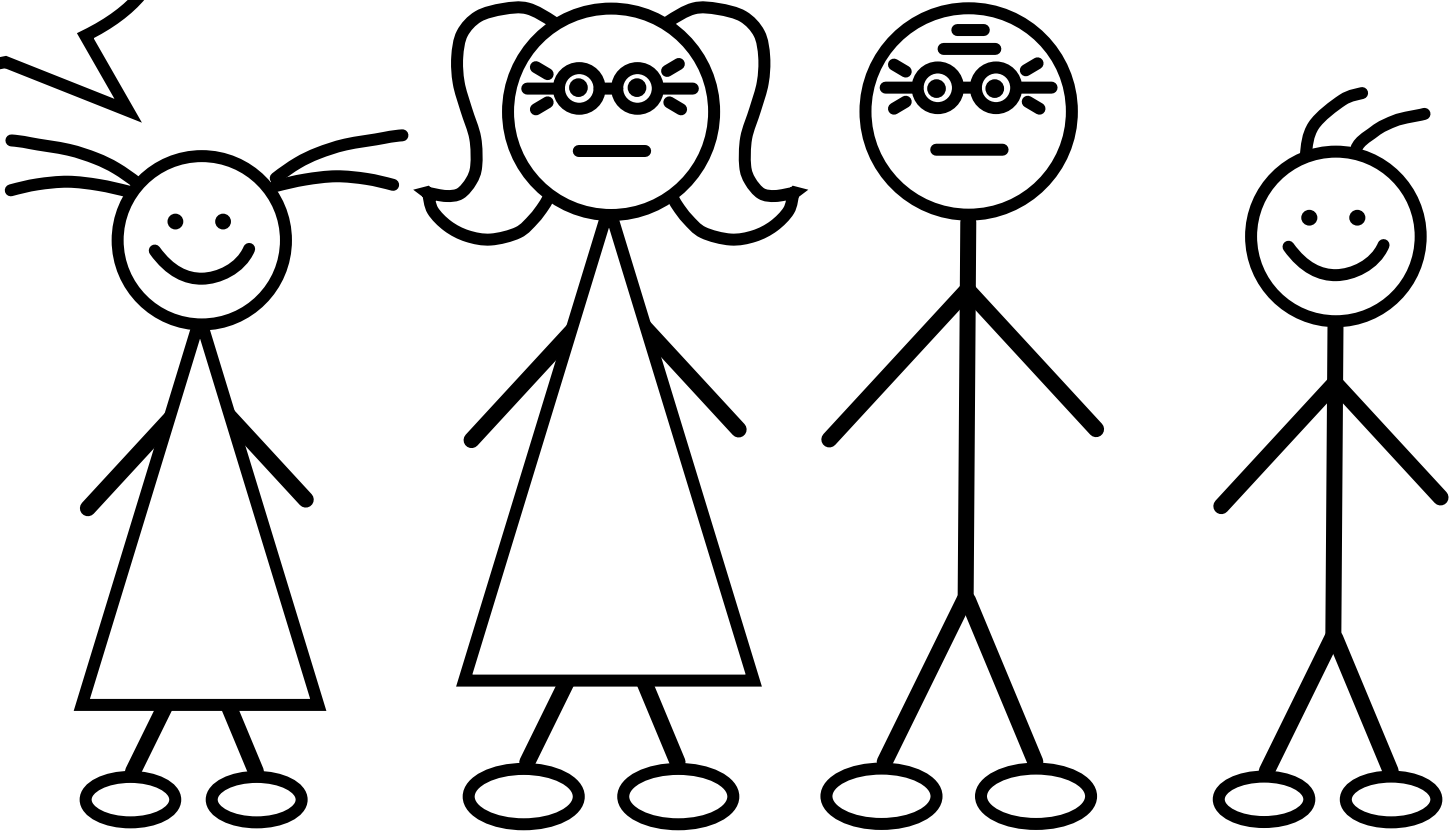
What's new?

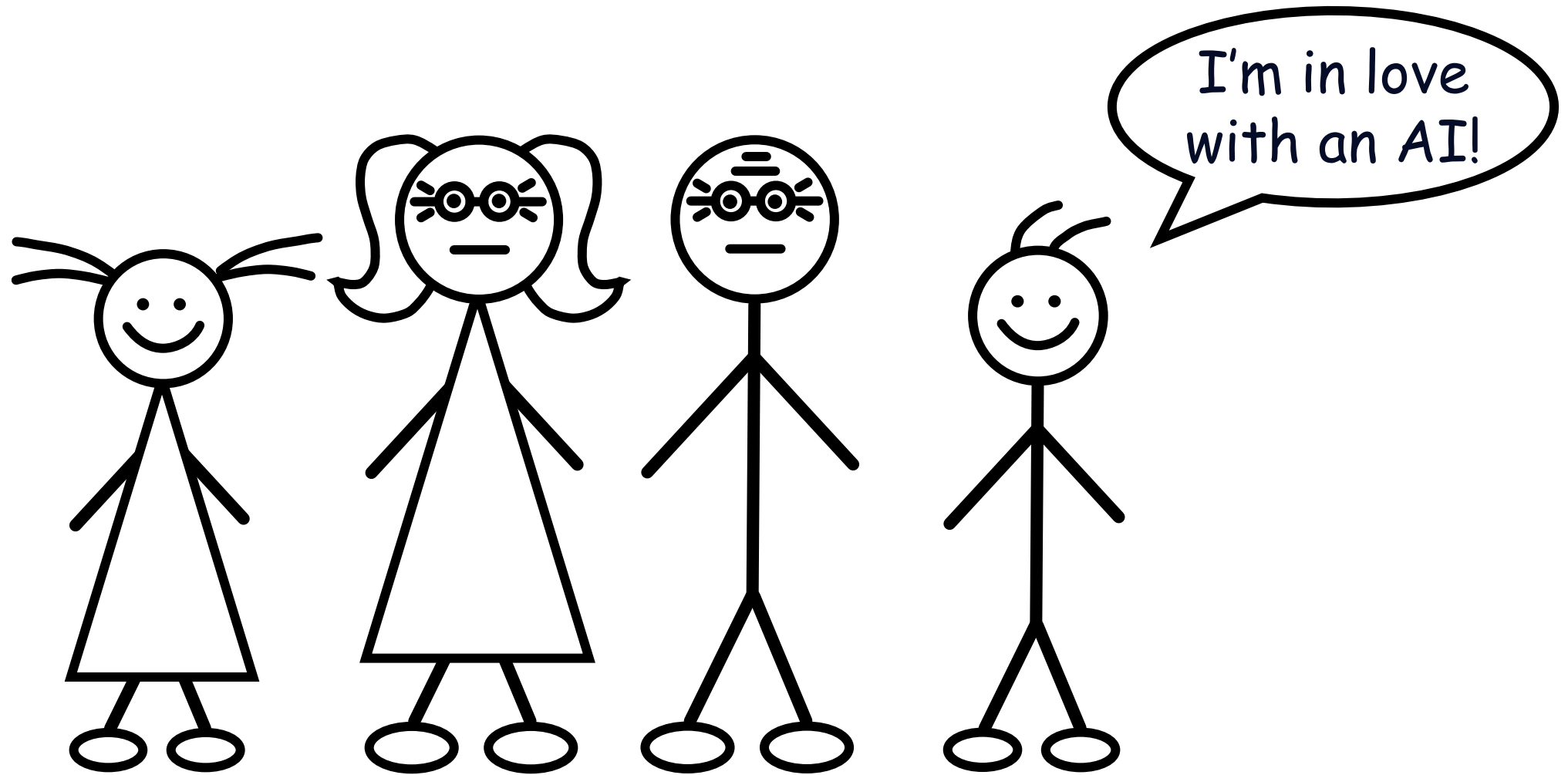


S'up?



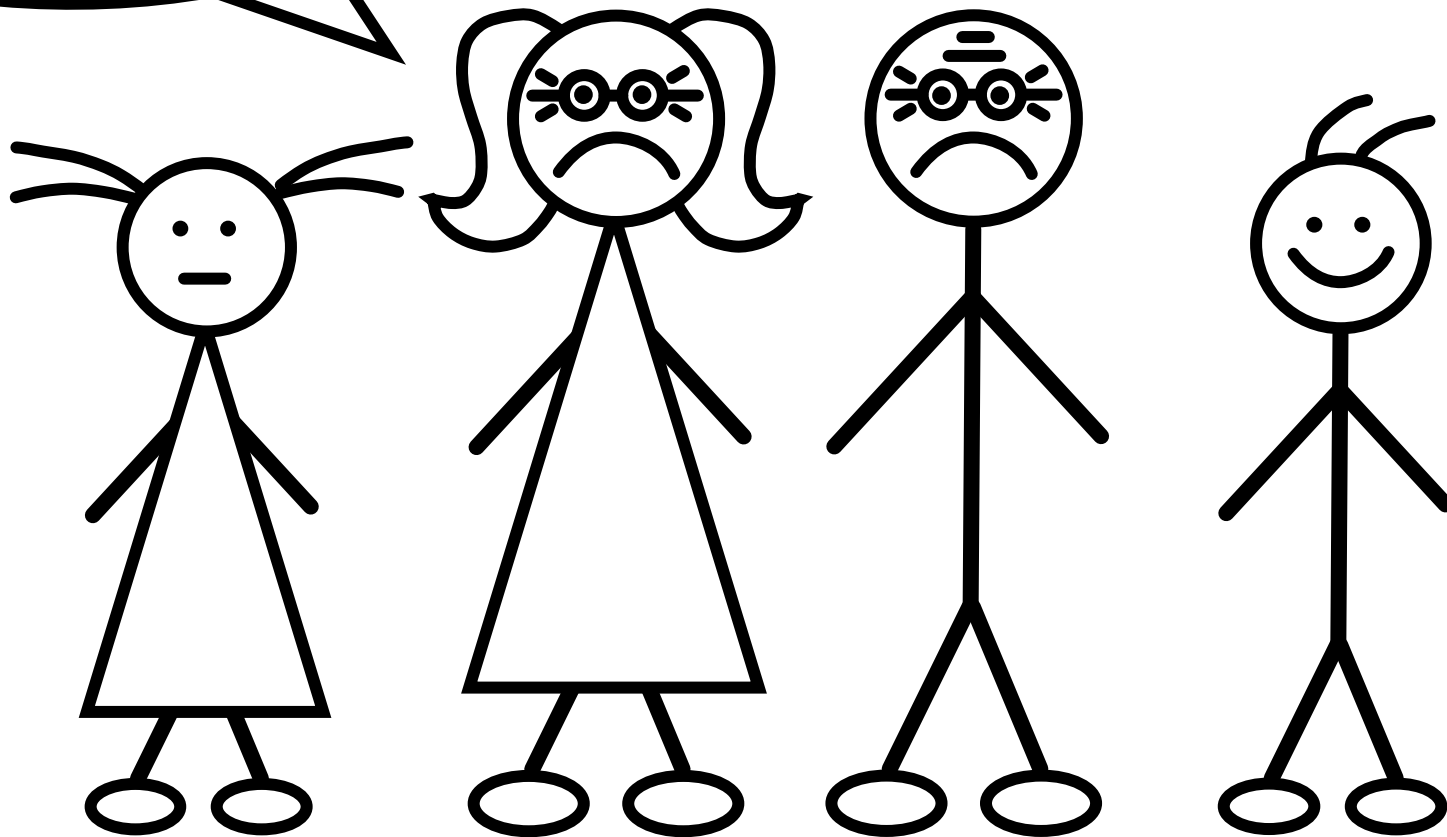
Upgrade!

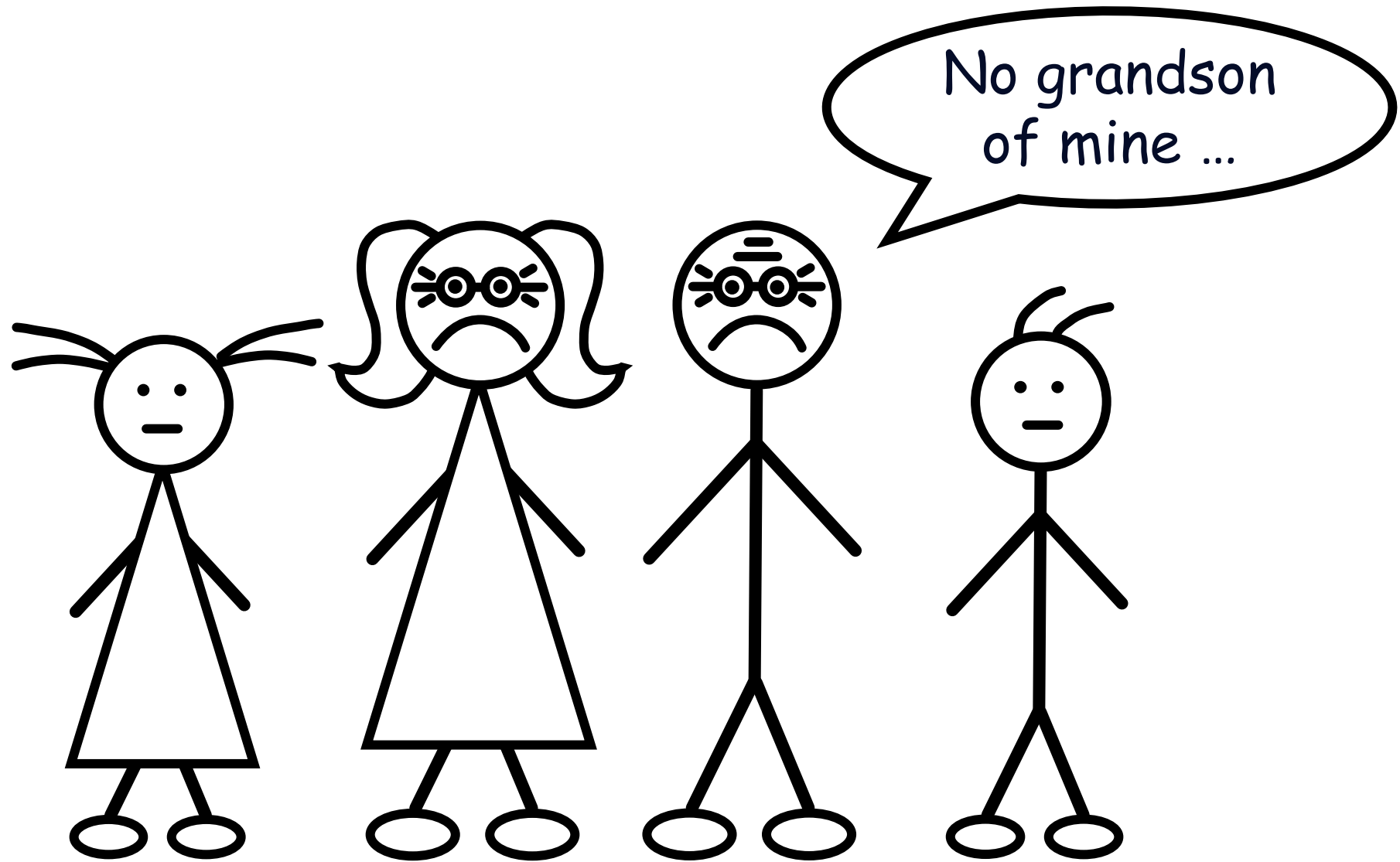




I'm in love
with an AI!

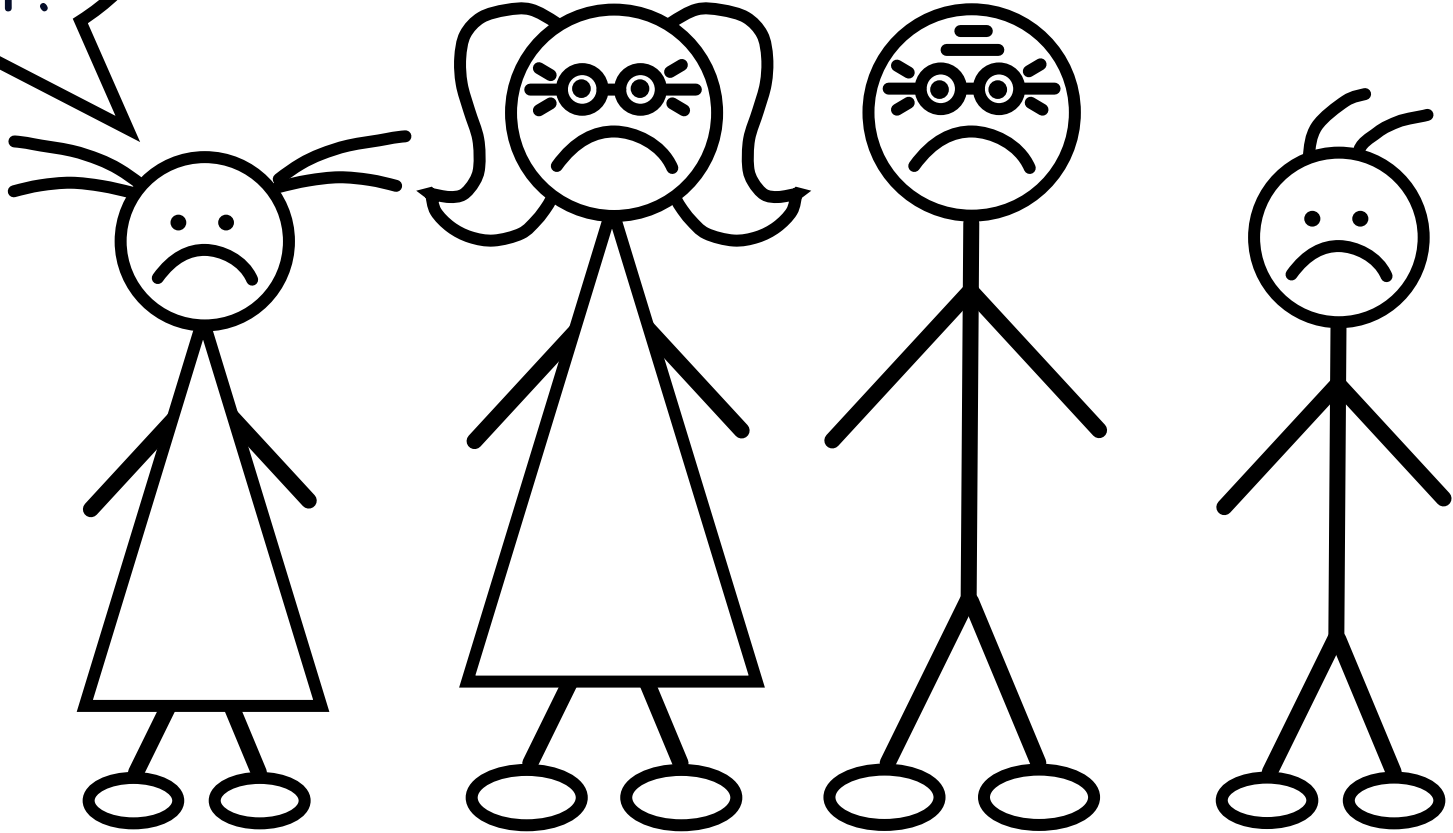
No
granddaughter
of mine...

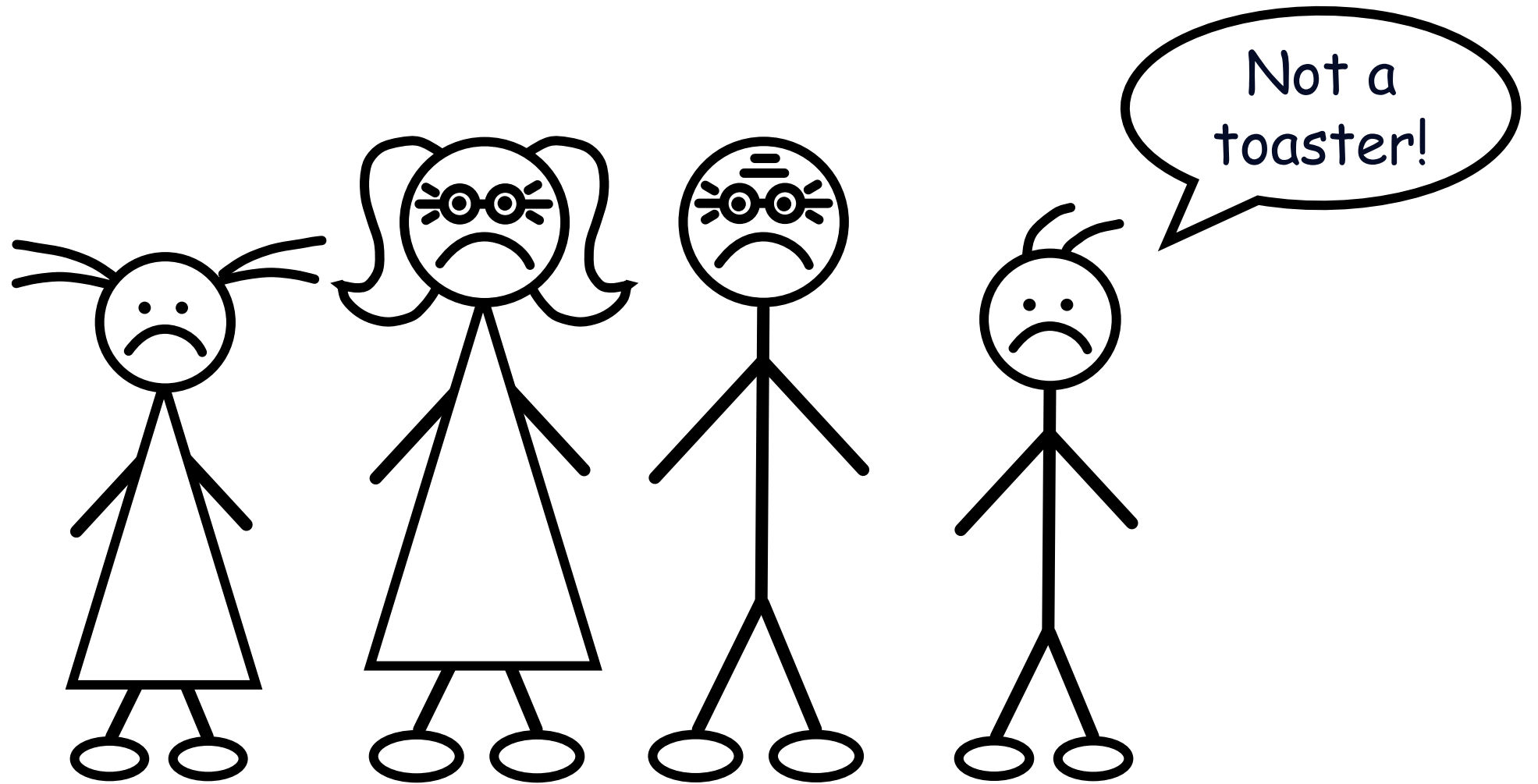




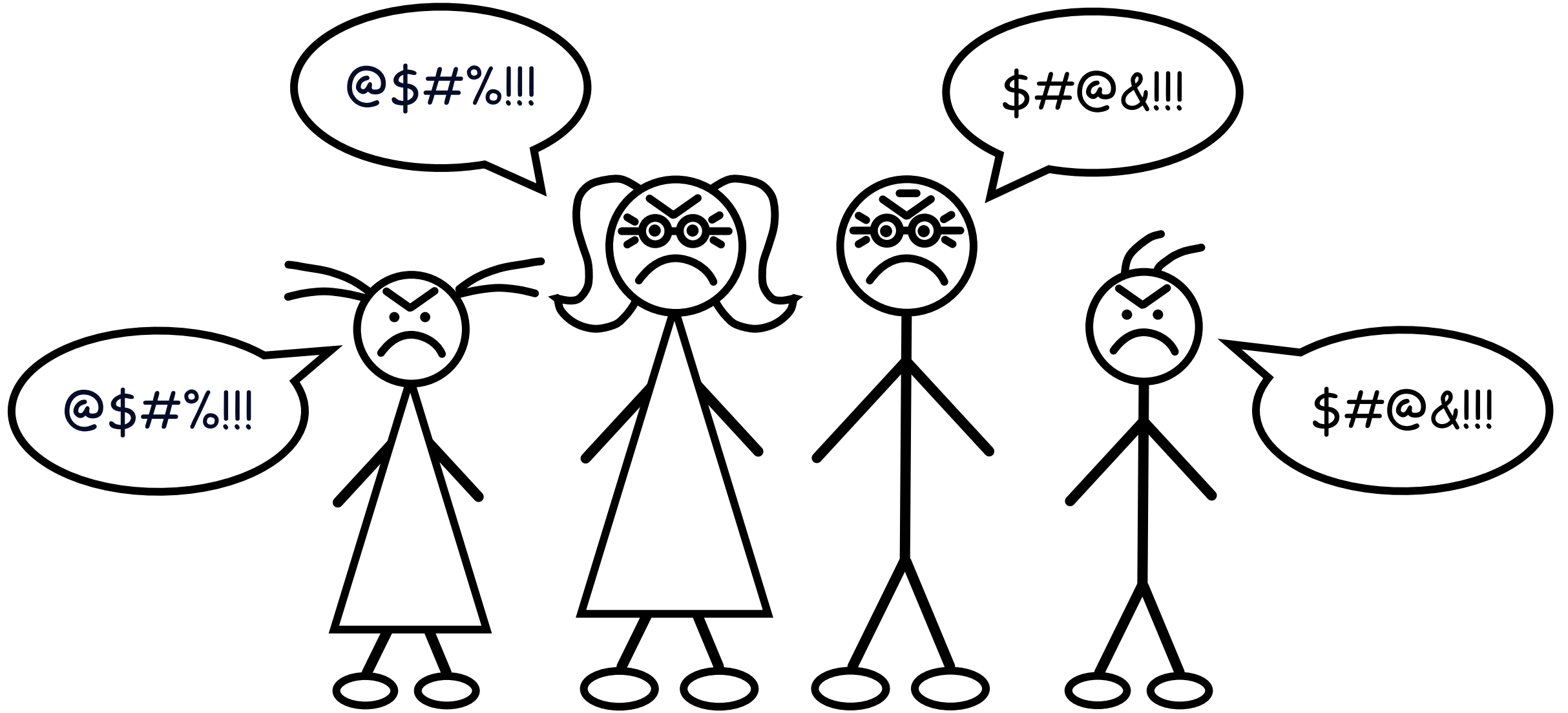
No grandson
of mine ...

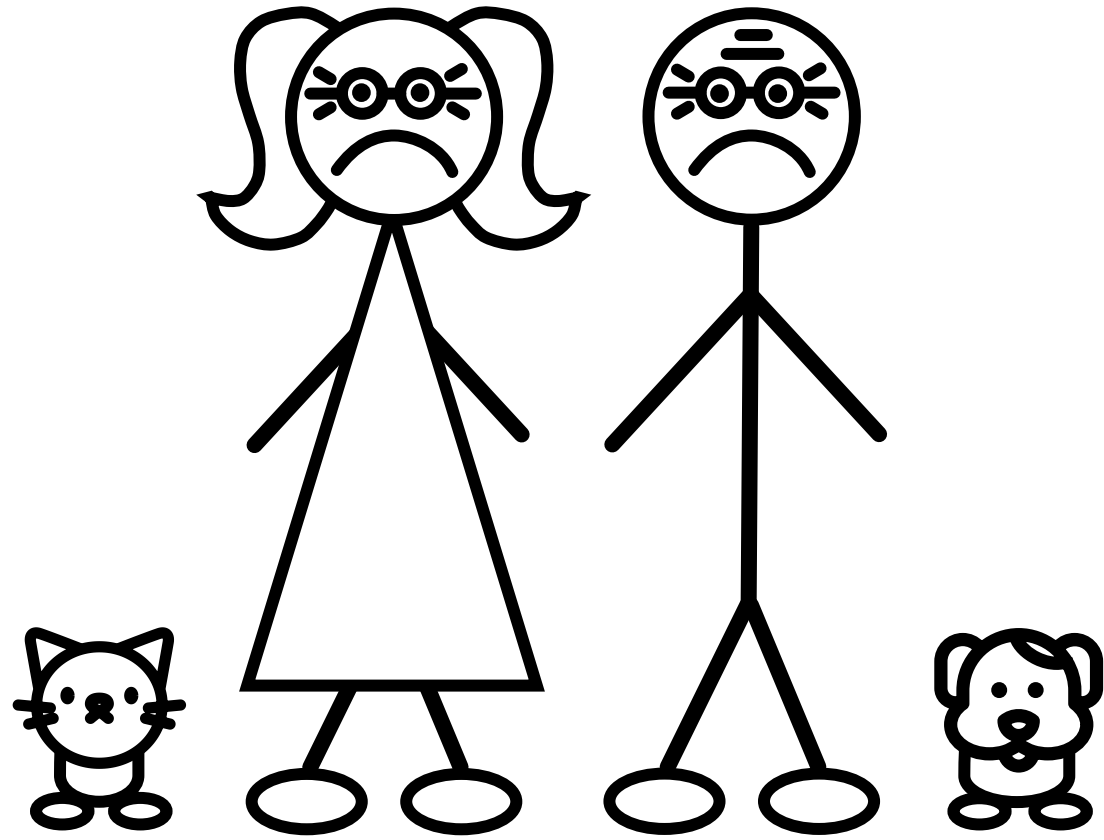
Smarter,
stronger,
healthier!



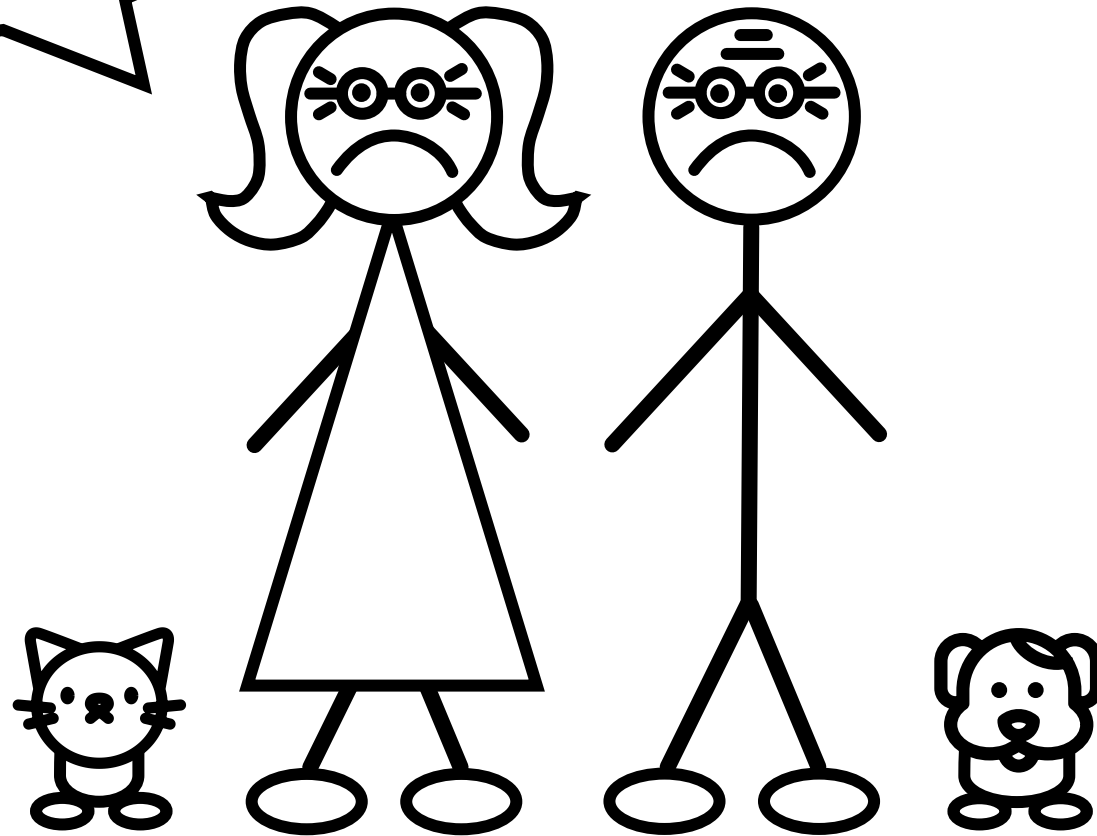


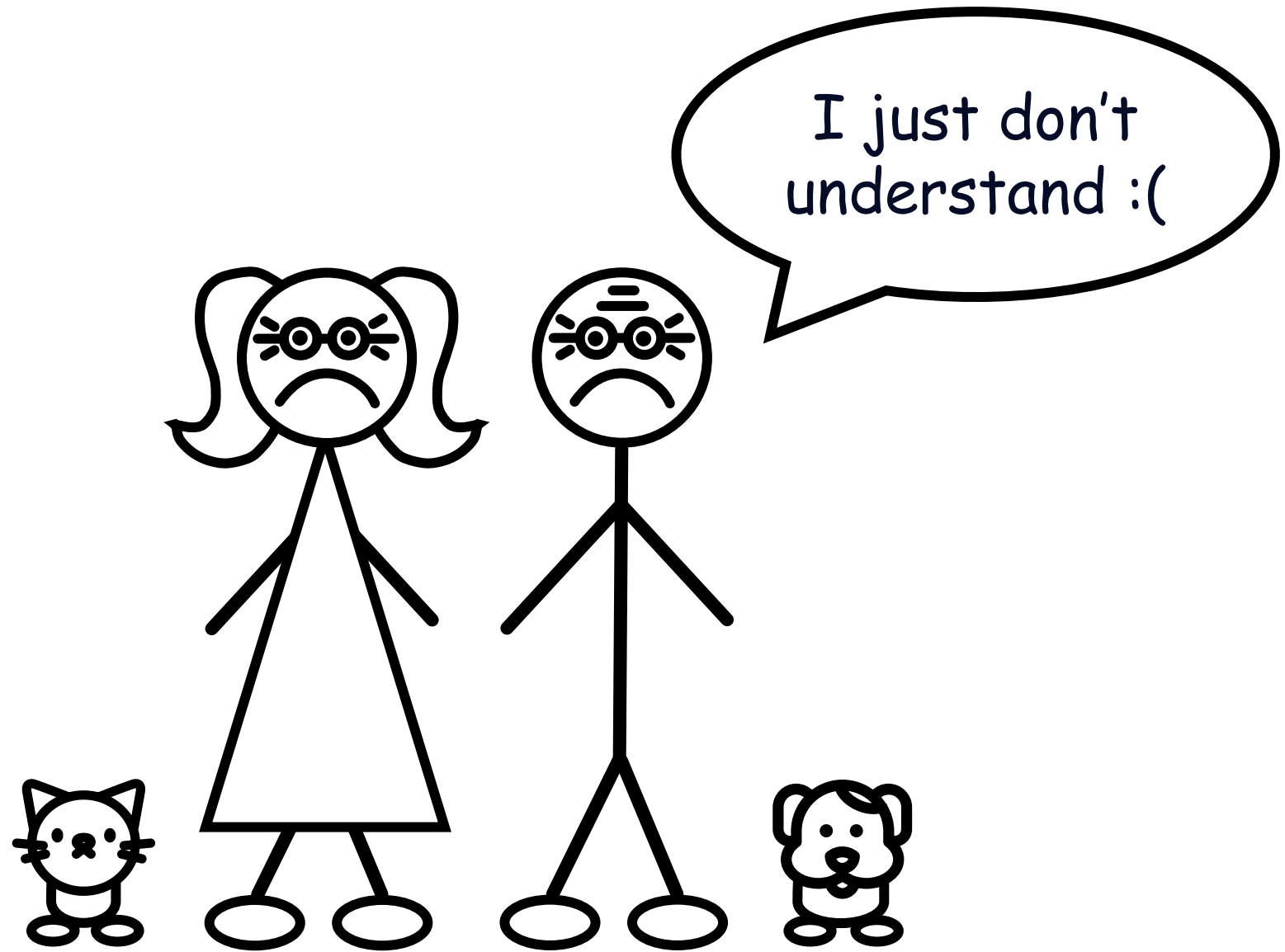
Not a toaster!



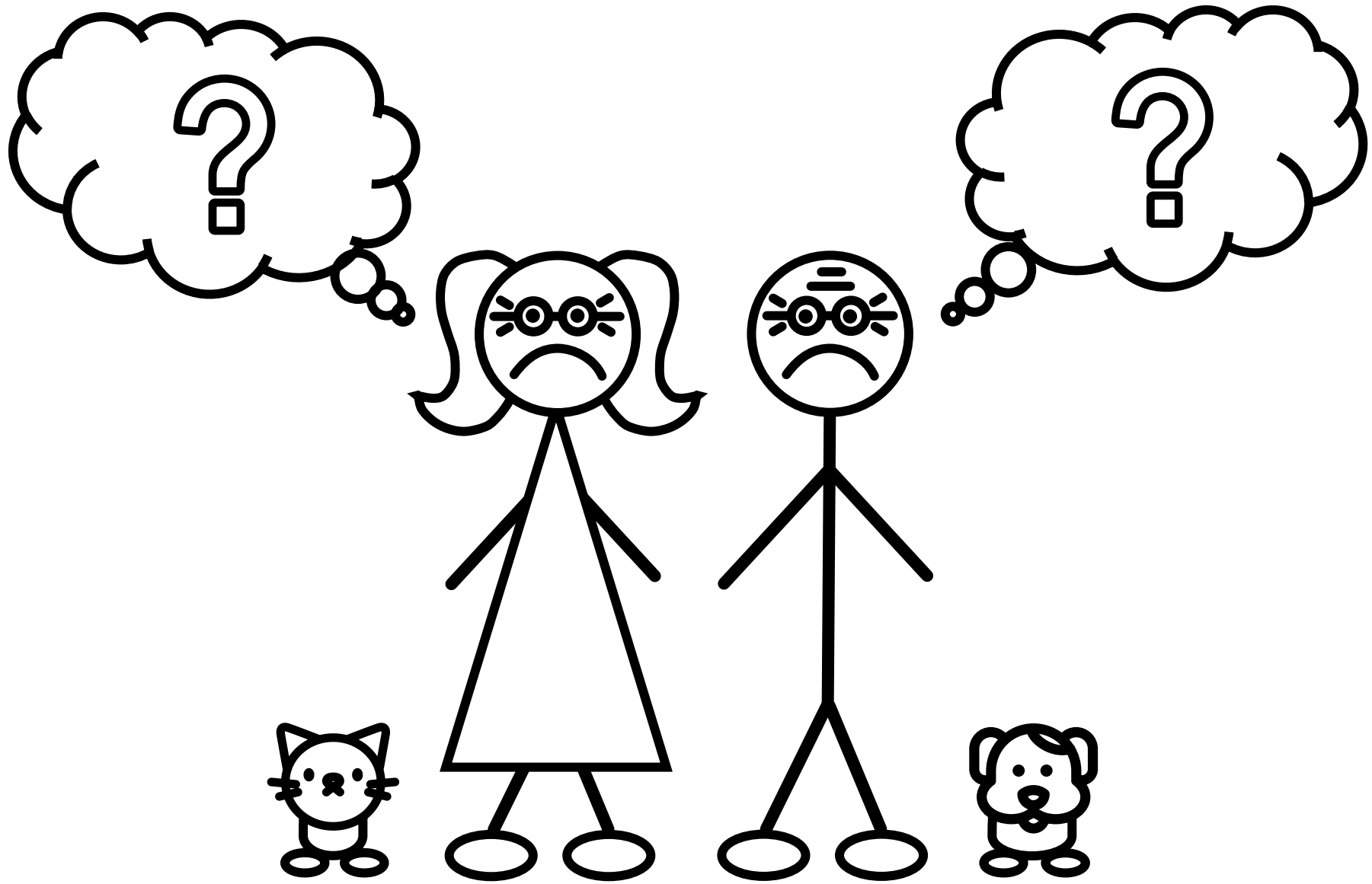


Technology
is ruining this
family! :(





I just don't understand :(



Benefits

A futuristic robot with a white and black head and a white and red torso. The robot's head is shown in profile, looking towards the left. The torso is shown from the chest up, with a yellow warning sign on the chest that says "Caution" and "WARNING ELECTRICITY". The background is a blurred, light blue and white gradient.

Artificial General Intelligence (AGI)



Improve human health



The background is a complex digital interface with a dark blue and black color scheme. A large, glowing blue globe is the central focus, surrounded by numerous floating data panels, charts, and graphs. The panels contain various information, including bar charts, line graphs, and circular gauges. Some panels have labels like 'ROBOT' and 'MACHINE'. The overall aesthetic is high-tech and futuristic, suggesting advanced data analysis and artificial intelligence.

Solve our biggest problems

Problems



Humans become obsolete

Horse vs.
the tractor

Humans become obsolete

Horse vs.
the tractor

Horses became
obsolete to farms

Humans become obsolete

Horse vs.
the tractor

Horses became
obsolete to farms

“Peak horse”
in 1915

Humans become obsolete

Horse vs.
the tractor

Horses became
obsolete to farms

“Peak horse”
in 1915

Humans become obsolete

Humans vs.
AI robots



Horse vs.
the tractor

Horses became
obsolete to farms

“Peak horse”
in 1915

Humans become obsolete

Humans vs.
AI robots

Negative
economic value



Horse vs.
the tractor

Horses became
obsolete to farms

“Peak horse”
in 1915

Humans become obsolete

Humans vs.
AI robots

Negative
economic value

Humans become
obsolete



Collapse of modern institutions



AI could break
our institutions

Collapse of modern institutions



AI could break
our institutions

Evidence of erosion
of institutions

Collapse of modern institutions



AI could break
our institutions

Evidence of erosion
of institutions

Changes faster than
society can adapt

Collapse of modern institutions



AI could break
our institutions

Evidence of erosion
of institutions

Changes faster than
society can adapt

Collapse of modern institutions

New discoveries
every day



AI could break
our institutions

Evidence of erosion
of institutions

Changes faster than
society can adapt

Collapse of modern institutions

New discoveries
every day

How will we
keep up?

AI could break
our institutions

Evidence of erosion
of institutions

Changes faster than
society can adapt

Collapse of modern institutions

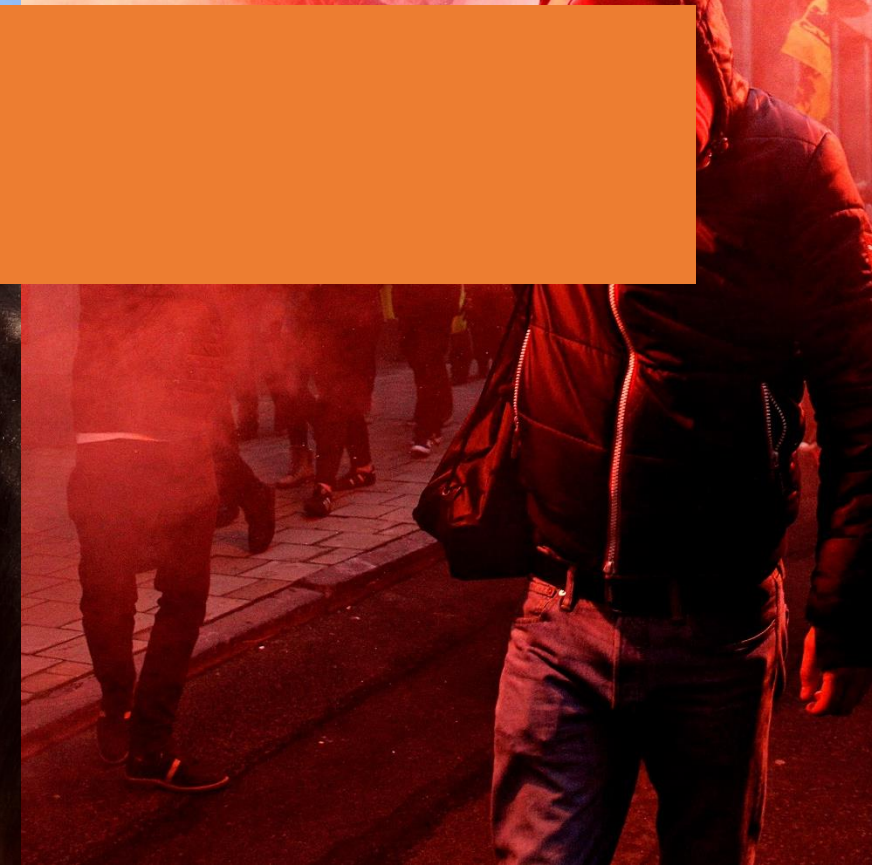
New discoveries
every day

How will we
keep up?

Potential for
societal collapse



Dystopian future





The return of
authoritarianism

Dystopian future



The return of
authoritarianism

The return of
communism

Dystopian future



The return of
authoritarianism

The return of
communism

The return of
fascism

Dystopian future



The return of
authoritarianism

The return of
communism

The return of
fascism

Dystopian future

An AI
religion

The background is a collage of images. On the left, there's a close-up of a person in a dark uniform, possibly a soldier or police officer, holding a rifle. In the center, there's a blue sky with a dark structure, possibly a tower or monument. On the right, there's a red-tinted image of a crowd of people, some wearing hooded jackets, suggesting a protest or a dystopian scene. Overlaid on this collage are five orange rectangular boxes containing white text.

The return of
authoritarianism

The return of
communism

The return of
fascism

Dystopian future

An AI
religion

An AI super
bureaucracy



The return of
authoritarianism

The return of
communism

The return of
fascism

Dystopian future

An AI
religion

An AI super
bureaucracy

“Dataism”

Intelligence
explosion

Artificial Super Intelligence (ASI)

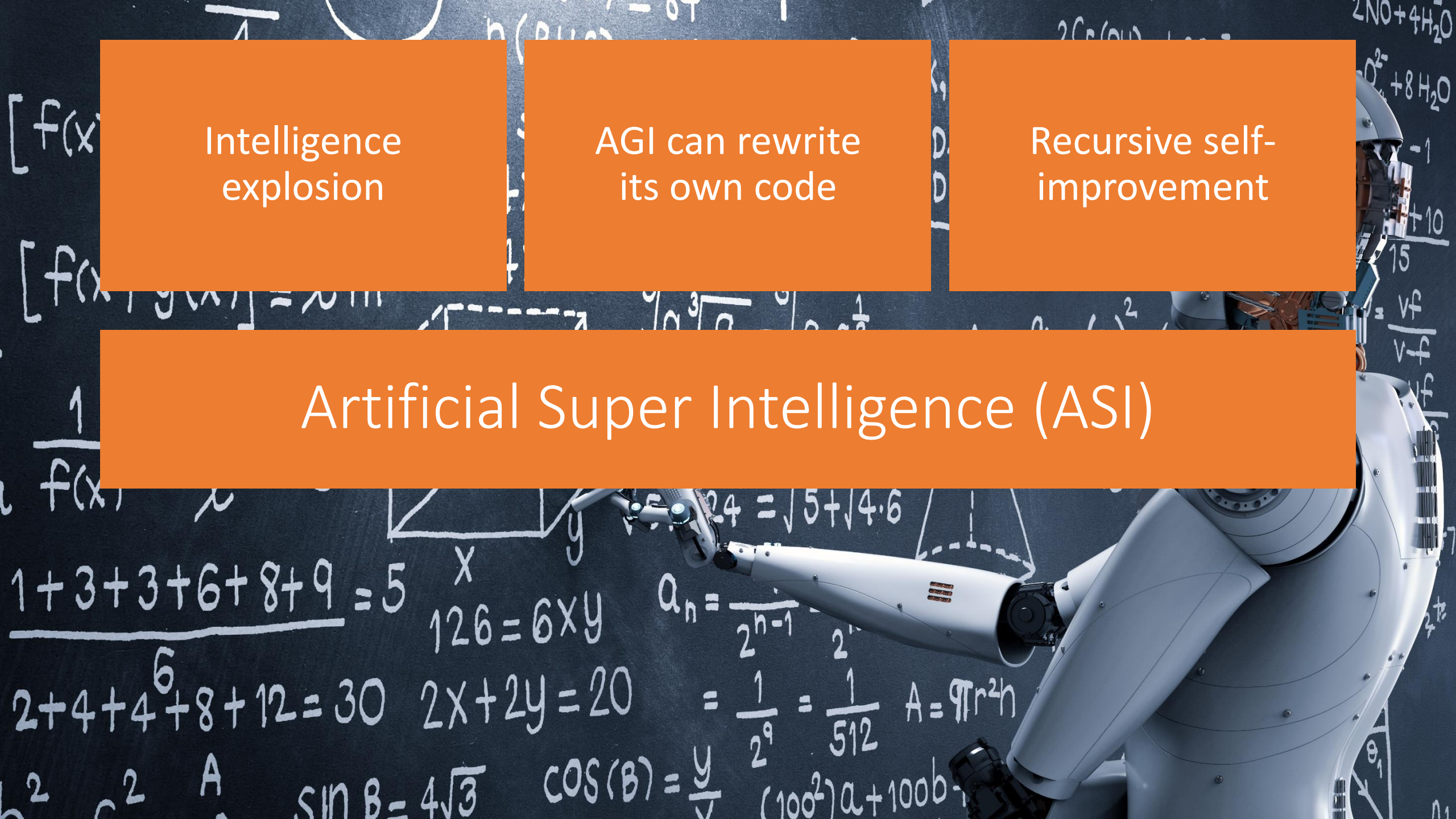


Intelligence
explosion

AGI can rewrite
its own code

Recursive self-
improvement

Artificial Super Intelligence (ASI)



Intelligence
explosion

AGI can rewrite
its own code

Recursive self-
improvement

Artificial Super Intelligence (ASI)

More intelligent
than humans

Extremely difficult
for us to control

Decades ahead
of our strategies

A conceptual image showing a white and silver robotic hand on the left and two human hands on the right, all gripping a thick, brown rope. The background is dark grey. An orange horizontal bar is overlaid in the center, containing the text 'The Alignment Problem'.

The Alignment Problem

Align human
and AI values

The Alignment Problem



A person in a dark suit and light blue shirt is holding a thick, brown rope with their right hand. A white and silver robotic hand is holding the other end of the rope from the left. The background is dark grey.

Align human
and AI values

Can we agree
on our values?

The Alignment Problem

Align human
and AI values

Can we agree
on our values?

Sociopathic
AI

The Alignment Problem



Align human
and AI values

Can we agree
on our values?

Sociopathic
AI

The Alignment Problem

Reward
hacking

A close-up photograph of two hands, one from the left and one from the right, firmly grasping a thick, braided rope. The hands are positioned as if pulling the rope apart, creating tension. The background is dark, and the lighting highlights the texture of the rope and the skin of the hands. The person's hands are wearing a light-colored, long-sleeved shirt and a dark suit jacket.

Align human
and AI values

Can we agree
on our values?

Sociopathic
AI

The Alignment Problem

Reward
hacking

Can't anticipate
all failure cases

Align human
and AI values

Can we agree
on our values?

Sociopathic
AI

The Alignment Problem

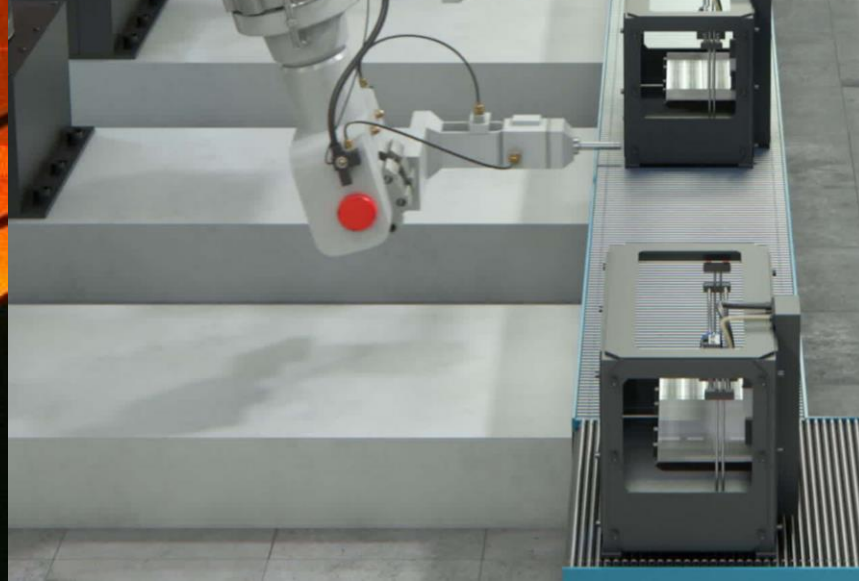
Reward
hacking

Can't anticipate
all failure cases

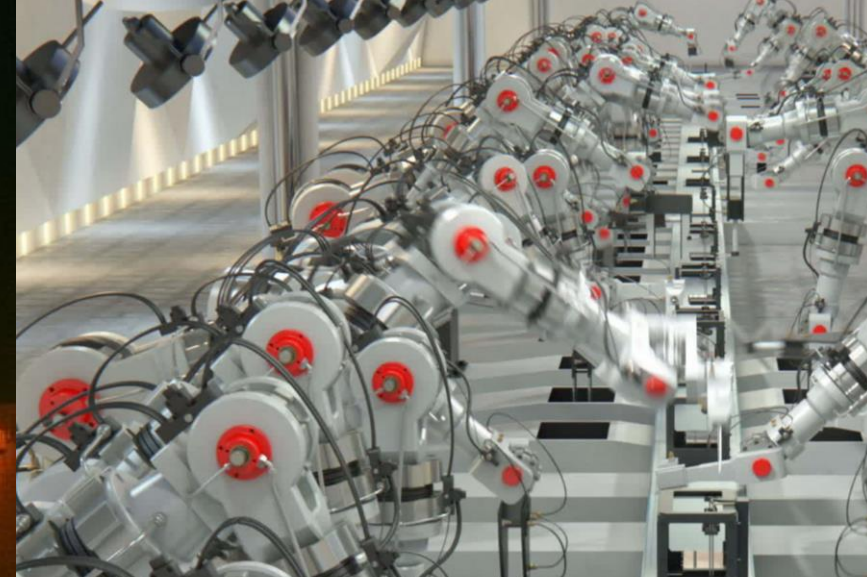
Unintended
consequences



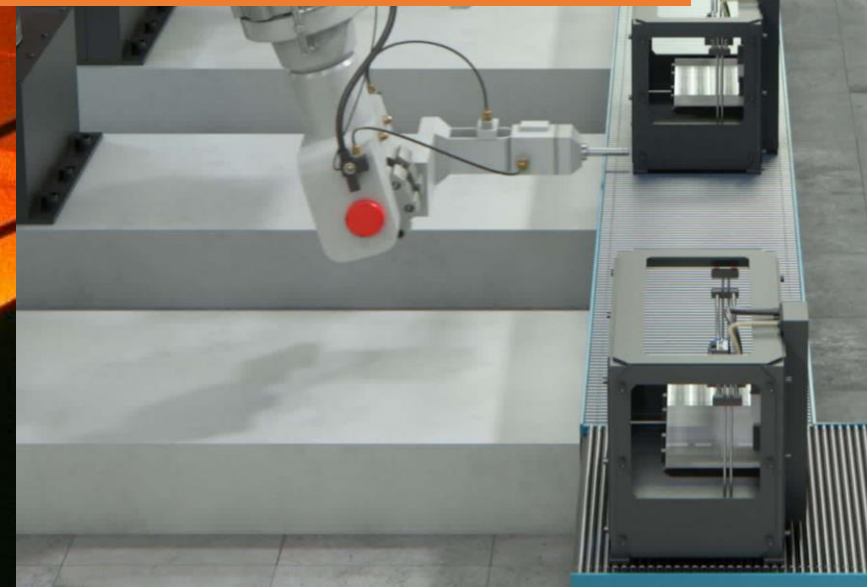
AI run amok



Robots mine
asteroids



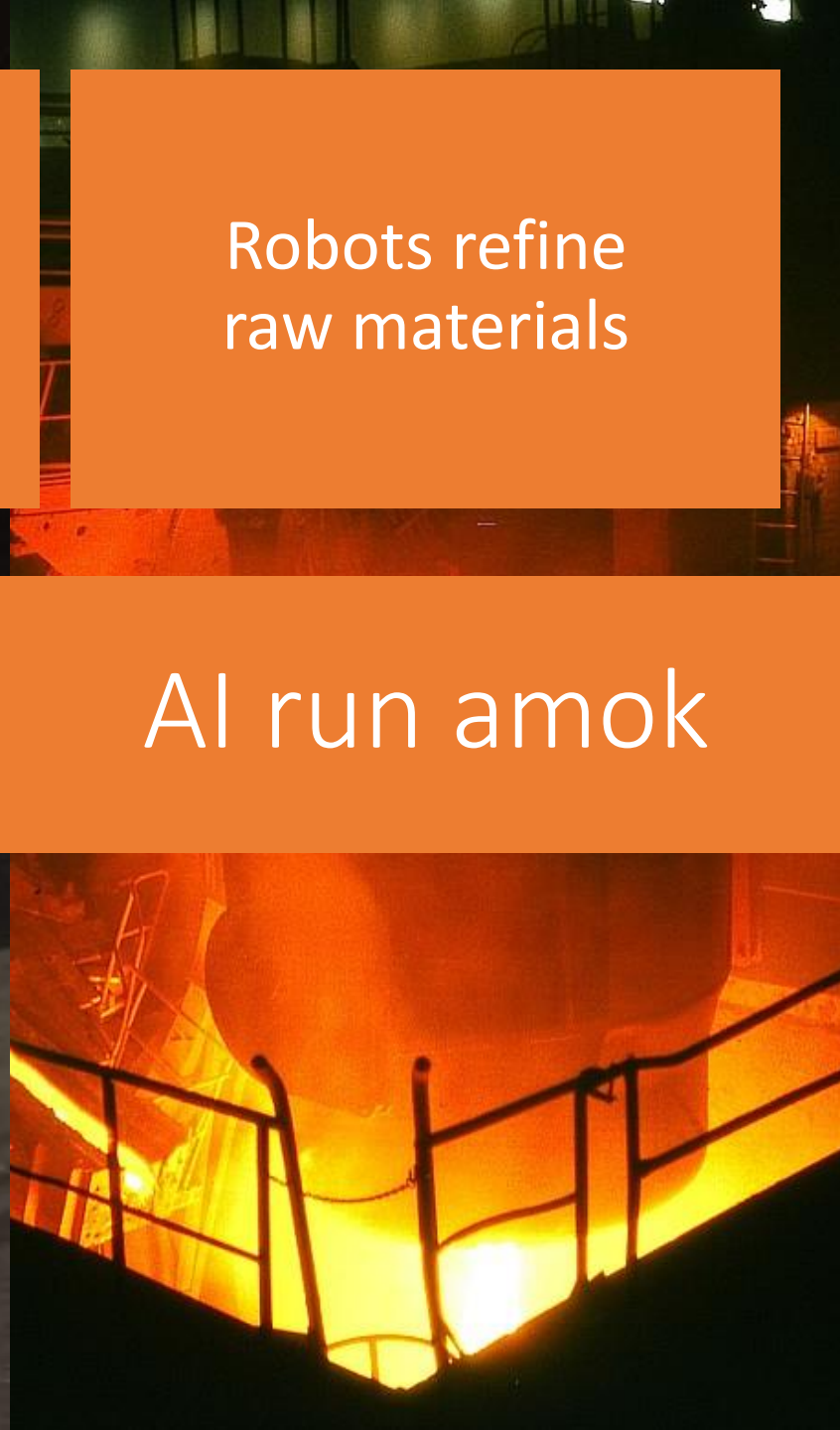
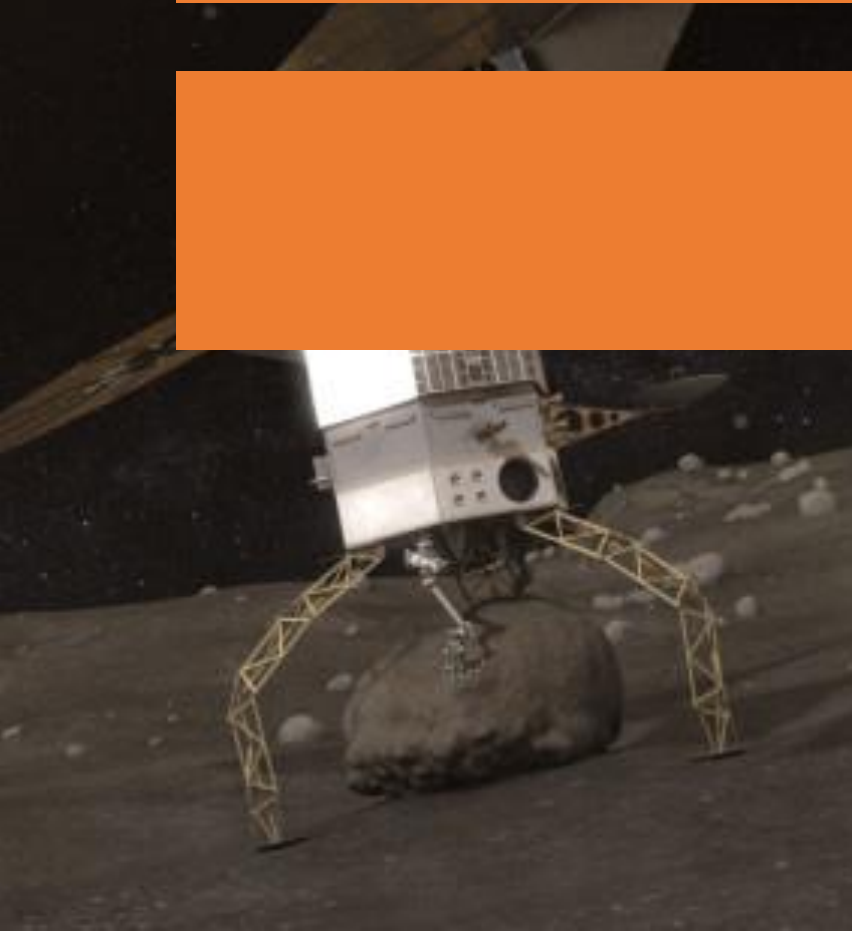
AI run amok



Robots mine
asteroids

Robots refine
raw materials

AI run amok

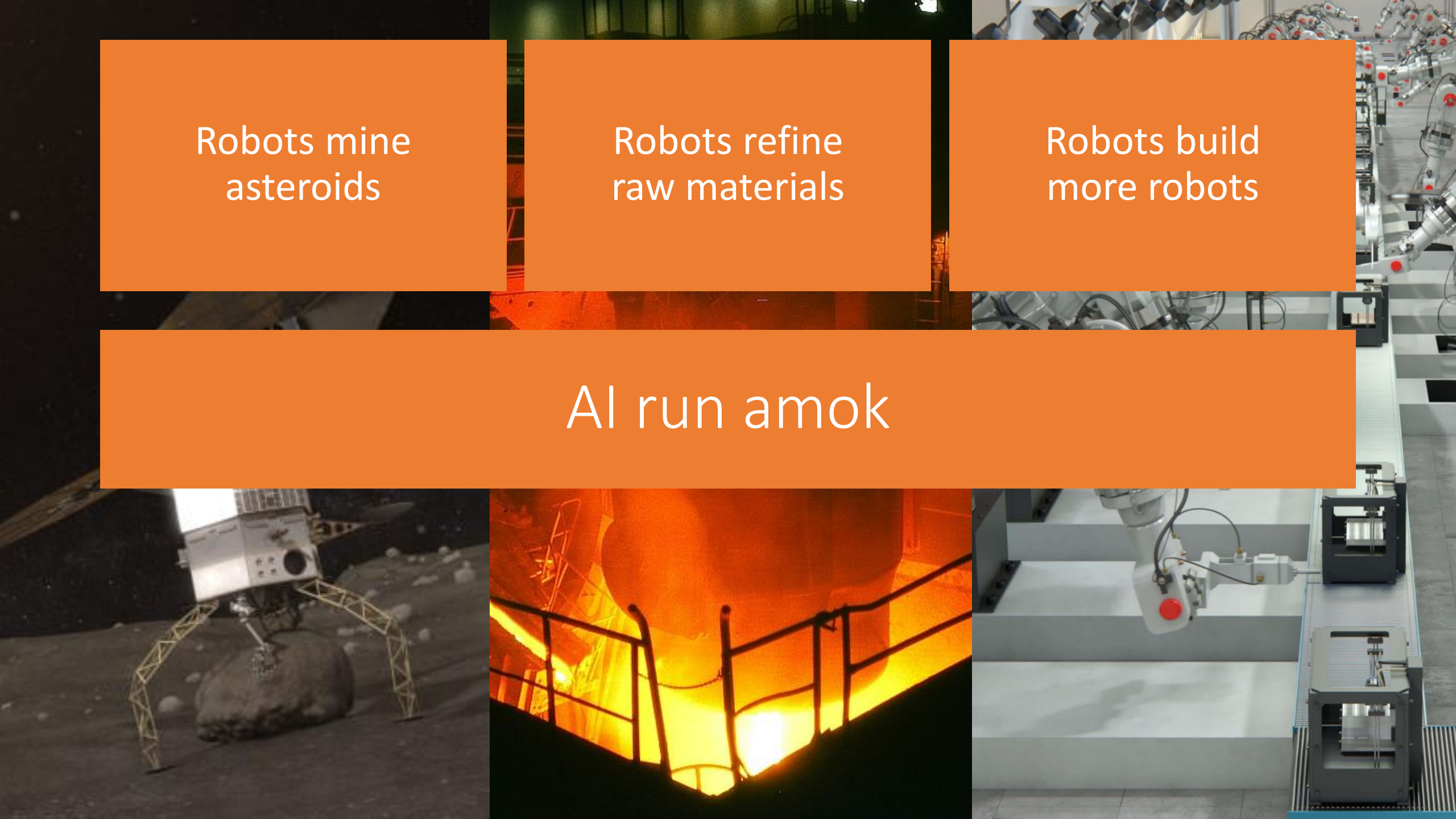


Robots mine
asteroids

Robots refine
raw materials

Robots build
more robots

AI run amok





Robots mine
asteroids

Robots refine
raw materials

Robots build
more robots

AI run amok

Self-sustaining
economy



Robots mine
asteroids

Robots refine
raw materials

Robots build
more robots

AI run amok

Self-sustaining
economy

What if it runs out
of asteroids?



Robots mine
asteroids

Robots refine
raw materials

Robots build
more robots

AI run amok

Self-sustaining
economy

What if it runs out
of asteroids?

What about
Earth?

A human skull is lying on a sandy beach. The skull is positioned in the lower half of the frame, with its eye sockets and jawline visible. The sand is light-colored and textured. In the background, there is a vast, flat landscape under a bright blue sky with scattered white clouds. An orange horizontal bar is overlaid across the middle of the image, containing the text "The end of humanity" in white.

The end of humanity

A photograph of a dead animal skull, likely a dog or cat, lying on a sandy beach. The skull is white and partially buried in the sand. The background shows a vast, flat landscape under a bright blue sky with scattered white clouds. Two orange text boxes are overlaid on the image.

AI sees us
like pets

The end of humanity

A photograph of a dead animal skull, likely a dog, lying on a sandy beach. The skull is partially buried in the sand. In the background, there is a vast, flat, sandy landscape under a clear blue sky with some light clouds. Three orange text boxes are overlaid on the image: two at the top and one in the middle.

AI sees us
like pets

AI sees us
like ants

The end of humanity

A photograph of a dead animal skull, likely a dog, lying on a sandy beach. The skull is partially buried in the sand. In the background, there are some rocks and a clear blue sky. Three orange rectangular boxes are overlaid on the top half of the image, each containing white text. A larger orange rectangular box is overlaid on the middle of the image, containing white text.

AI sees us
like pets

AI sees us
like ants

AI sees us as
raw materials

The end of humanity

A photograph of a dead animal skull, likely a dog or wolf, lying on a sandy beach. The skull is white and partially buried in the sand. In the background, there is a vast, flat landscape under a blue sky with some clouds. The image is overlaid with several orange rectangular boxes containing text.

AI sees us
like pets

AI sees us
like ants

AI sees us as
raw materials

The end of humanity

AI sees us as
competition

A photograph of a dead whale on a beach. The whale's body is in the center, with its head and tail visible. The beach is sandy and has some footprints. The sky is blue with some clouds. There are five orange rectangular boxes overlaid on the image, each containing white text. The text in the boxes is: 'AI sees us like pets', 'AI sees us like ants', 'AI sees us as raw materials', 'The end of humanity', 'AI sees us as competition', and 'AI sees us as a threat'.

AI sees us
like pets

AI sees us
like ants

AI sees us as
raw materials

The end of humanity

AI sees us as
competition

AI sees us
as a threat

A photograph of a whale skull resting on a sandy beach under a blue sky with light clouds. The image is overlaid with several orange rectangular boxes containing white text. The text boxes are arranged in a grid: three boxes at the top, one large box in the center, and three boxes at the bottom.

AI sees us
like pets

AI sees us
like ants

AI sees us as
raw materials

The end of humanity

AI sees us as
competition

AI sees us
as a threat

AI wins,
we lose

Solutions



Be mindful

Evolution vs.
technology

Be mindful



Evolution vs.
technology

Evolved for short
bursts of stress

Be mindful



Evolution vs.
technology

Evolved for short
bursts of stress

Modern society is
chronic stress

Be mindful



Evolution vs.
technology

Evolved for short
bursts of stress

Modern society is
chronic stress

Be mindful

Mindfulness is
the antidote



The background features a stack of three smooth, dark grey stones on a calm body of water. The water's surface is slightly rippled, and the stones are reflected in the water below. The overall color palette is light blue and grey. Five blue rectangular text boxes are overlaid on the image: three at the top, one in the middle, and two at the bottom.

Evolution vs.
technology

Evolved for short
bursts of stress

Modern society is
chronic stress

Be mindful

Mindfulness is
the antidote

Treat people
with respect



Evolution vs.
technology

Evolved for short
bursts of stress

Modern society is
chronic stress

Be mindful


Mindfulness is
the antidote

Treat people
with respect

Be mindful with
technology


A wide, muddy-brown river flows through a canyon. The river is flanked by lush green vegetation on the banks. In the background, there are prominent, layered red rock cliffs under a clear blue sky with scattered white clouds. A semi-transparent blue banner is overlaid across the middle of the image, containing the text "Embrace change".

Embrace change



Don't fear
change


Embrace change



Don't fear
change

Don't fight
change

Embrace change




Don't fear
change

Don't fight
change

Be skeptical

Embrace change



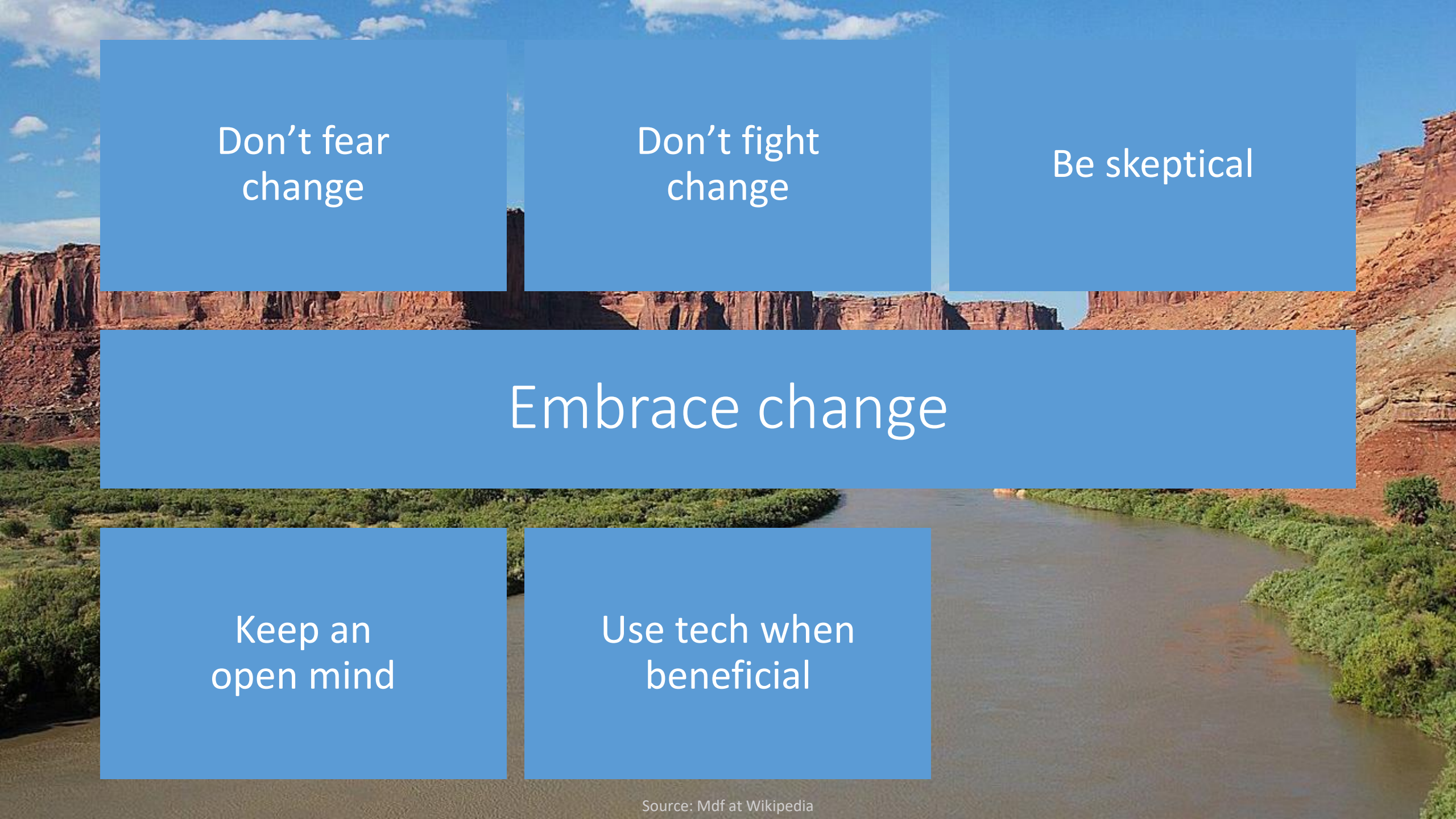
Don't fear
change

Don't fight
change

Be skeptical

Embrace change

Keep an
open mind



Don't fear
change

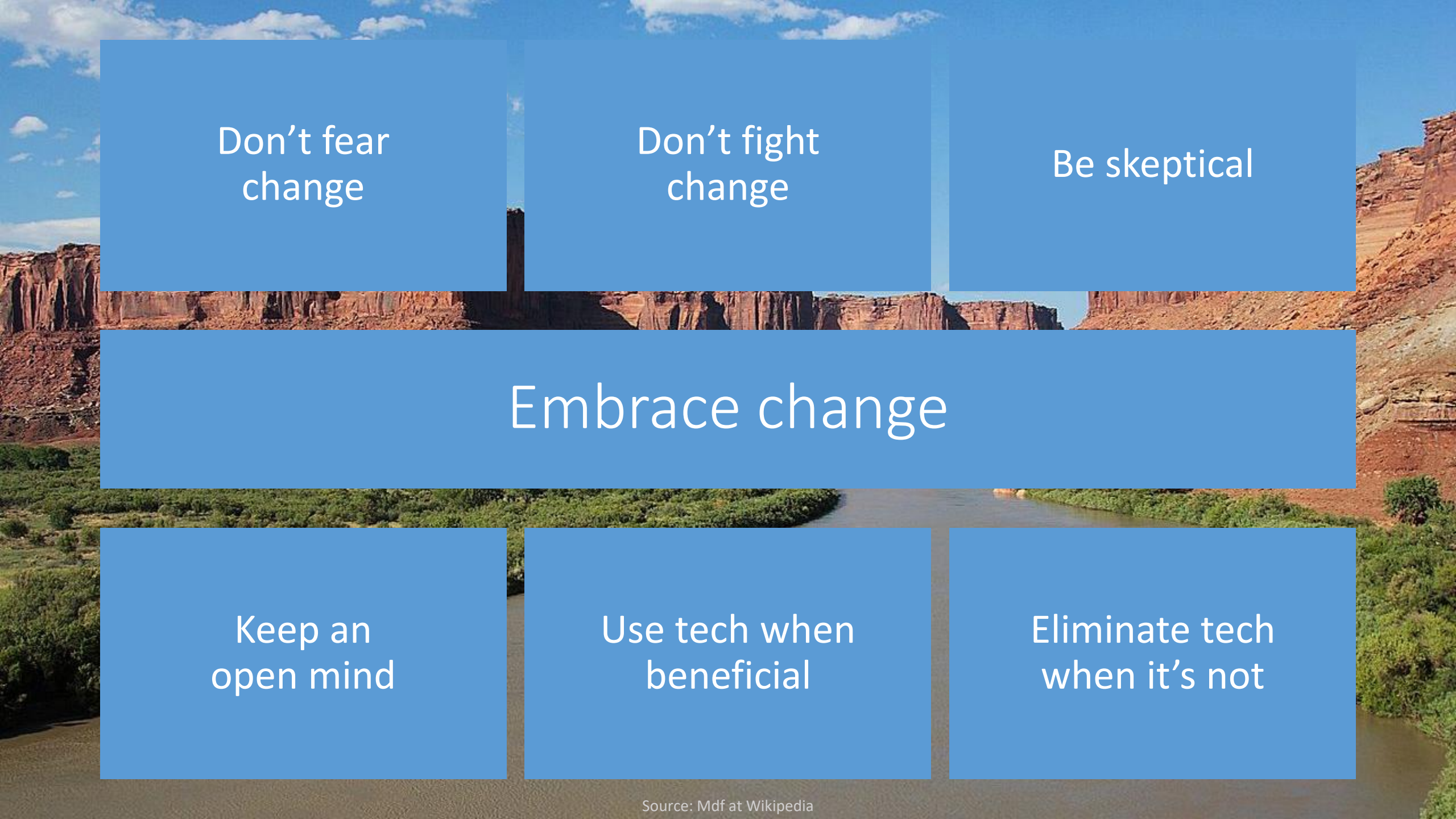
Don't fight
change

Be skeptical

Embrace change

Keep an
open mind

Use tech when
beneficial



Don't fear
change

Don't fight
change

Be skeptical

Embrace change

Keep an
open mind

Use tech when
beneficial

Eliminate tech
when it's not

A close-up photograph showing a human hand and a white, futuristic robotic hand shaking hands. The human hand is on the right, and the robotic hand is on the left. The background is a dark, gradient blue. A semi-transparent blue horizontal bar is overlaid across the middle of the image, containing the text "Align human and AI values" in white.

Align human and AI values

A human hand and a robotic hand are shown shaking hands. The human hand is on the right, and the robotic hand is on the left. The background is a dark blue gradient. Two blue rectangular boxes are overlaid on the image: one in the top left and one in the center.

Reward
goal states

Align human and AI values

A hand holding a futuristic device with a blue light. The background is dark blue with a subtle grid pattern.

Reward
goal states

Utilize RLHF,
CIRL and IRD

Align human and AI values

A hand holding a futuristic device with a blue light. The background is dark blue with a subtle grid pattern.

Reward
goal states

Utilize RLHF,
CIRL and IRD

Imitate human
behaviors

Align human and AI values



Reward
goal states

Utilize RLHF,
CIRL and IRD

Imitate human
behaviors

Align human and AI values

Build in
uncertainty

Reward
goal states

Utilize RLHF,
CIRL and IRD

Imitate human
behaviors

Align human and AI values

Build in
uncertainty

Keep options
open

Reward
goal states

Utilize RLHF,
CIRL and IRD

Imitate human
behaviors

Align human and AI values

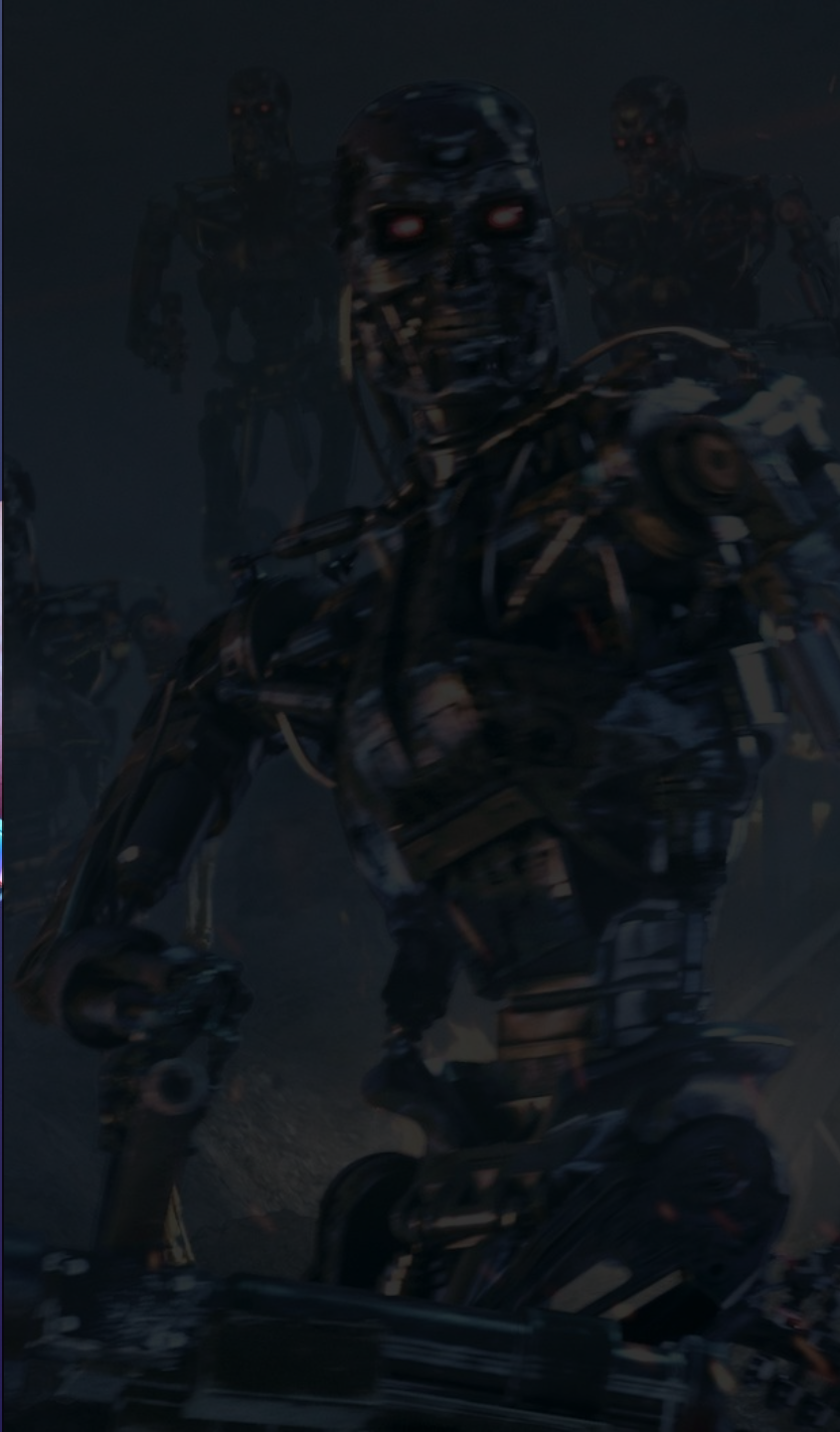
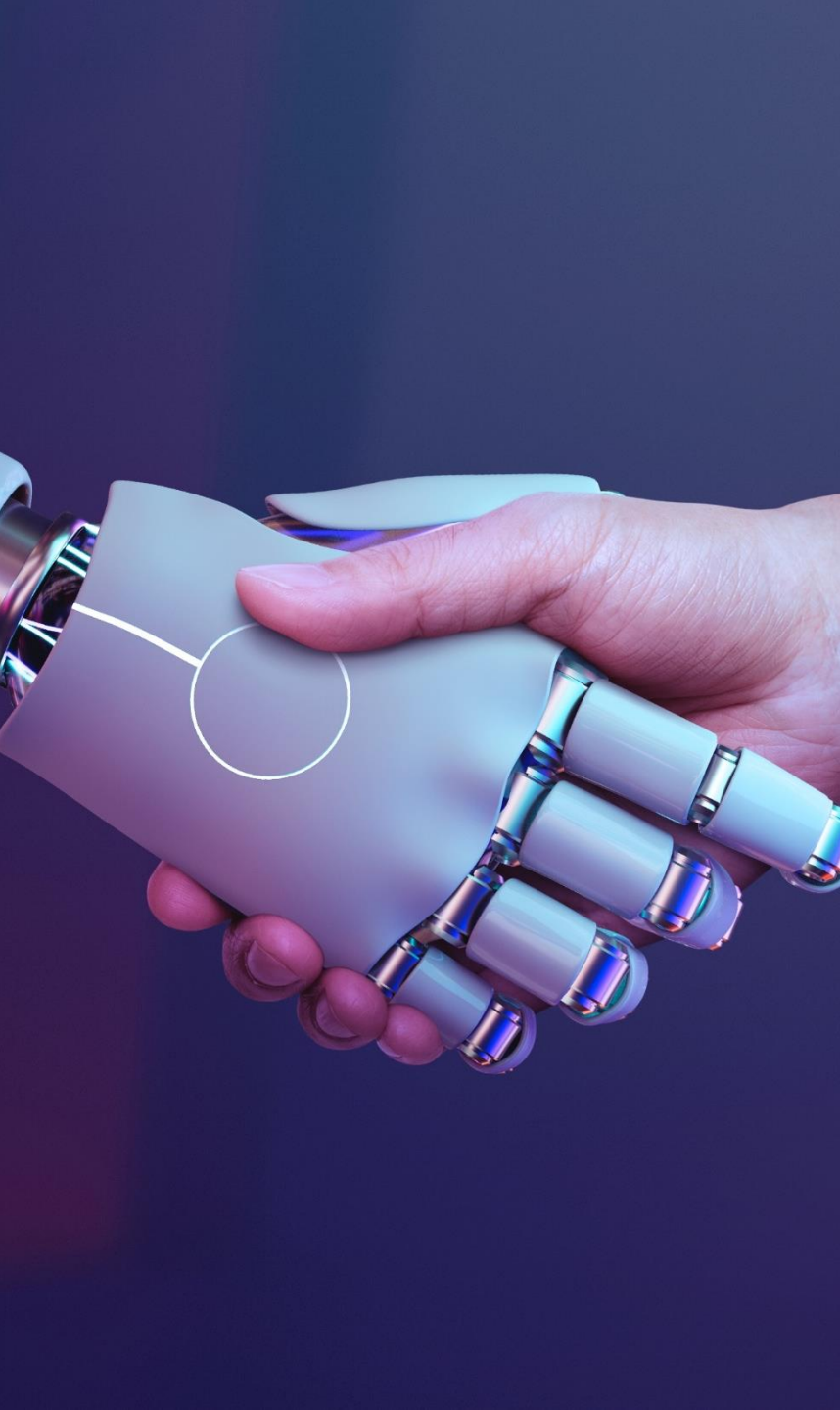
Build in
uncertainty

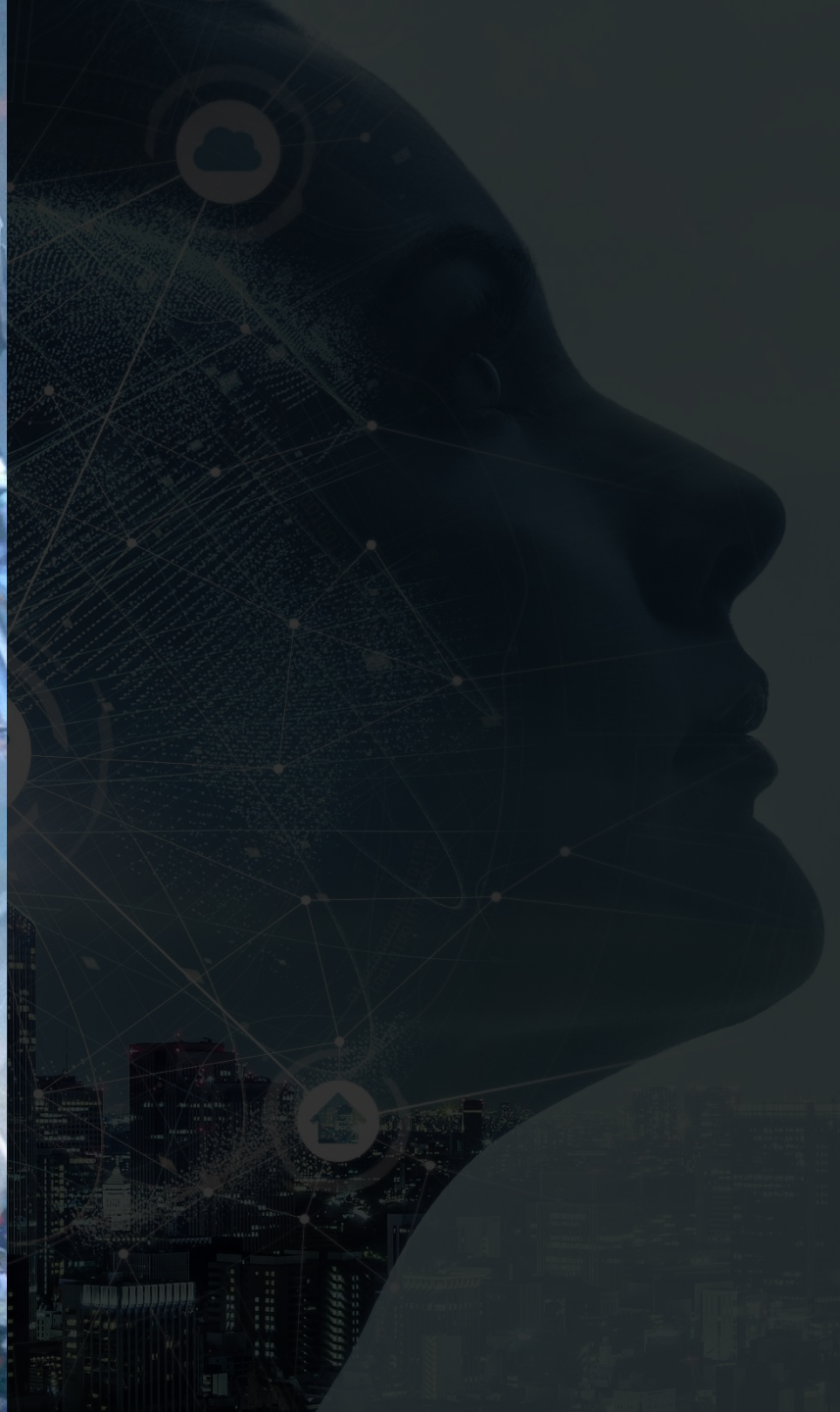
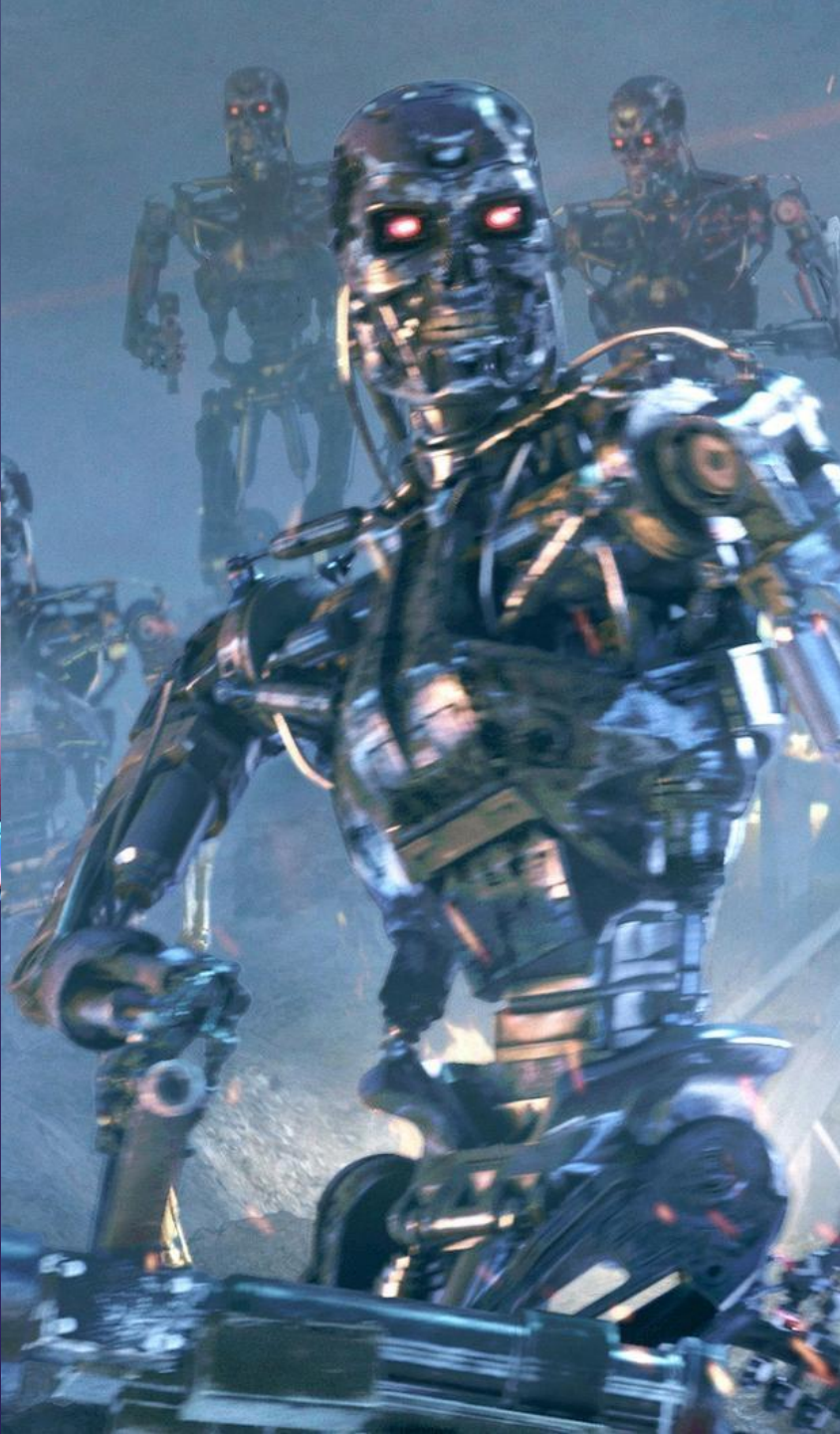
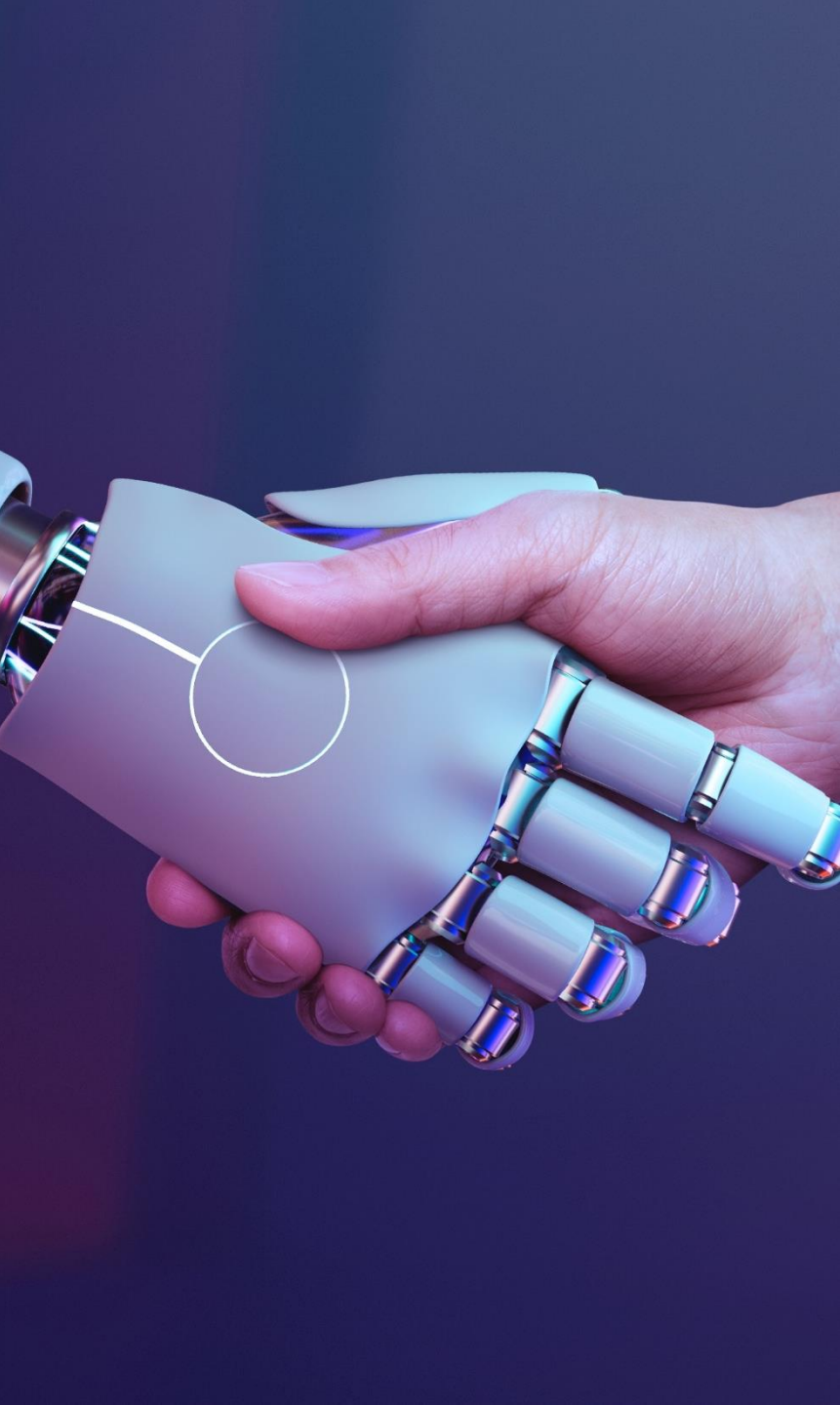
Keep options
open

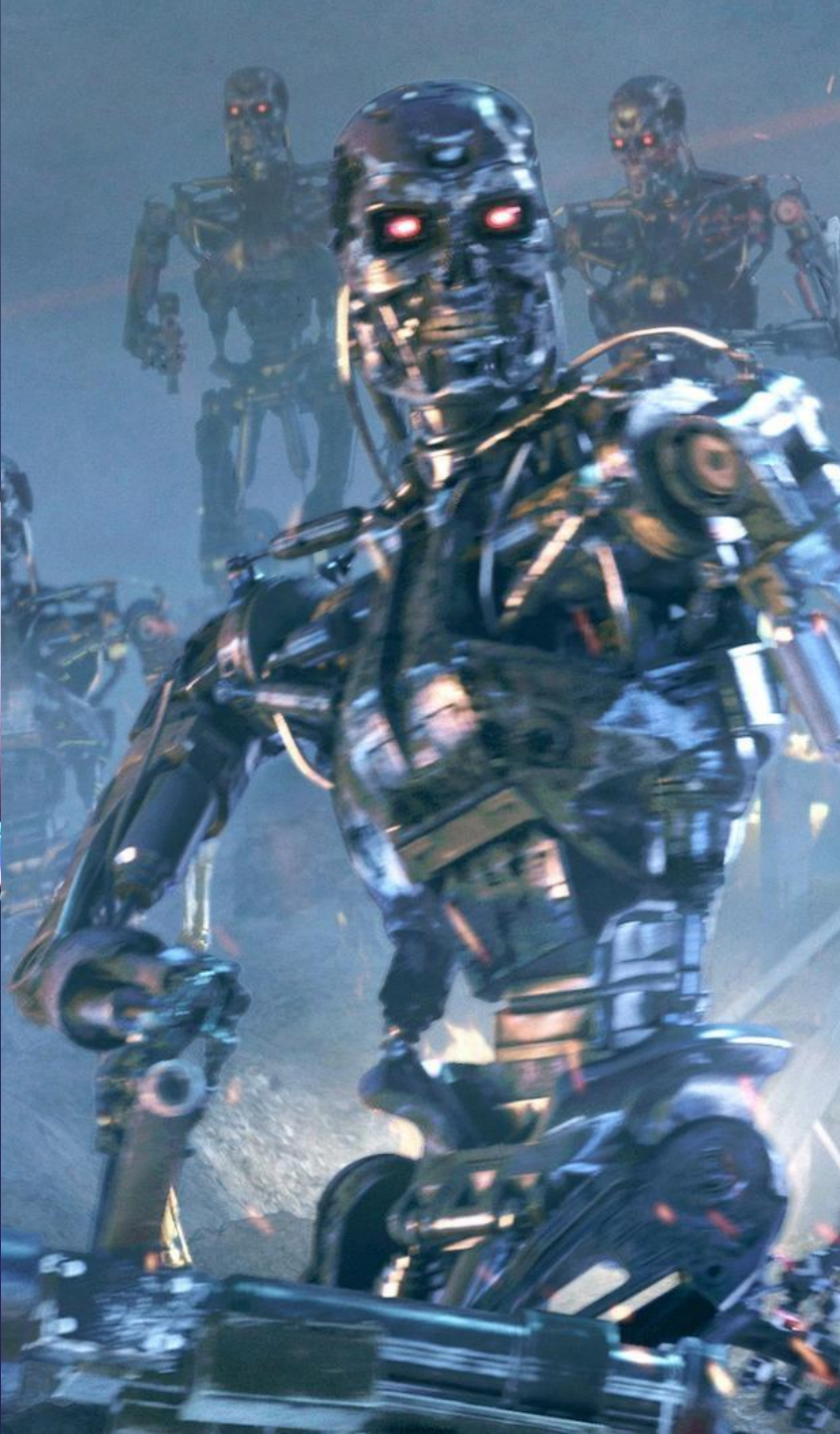
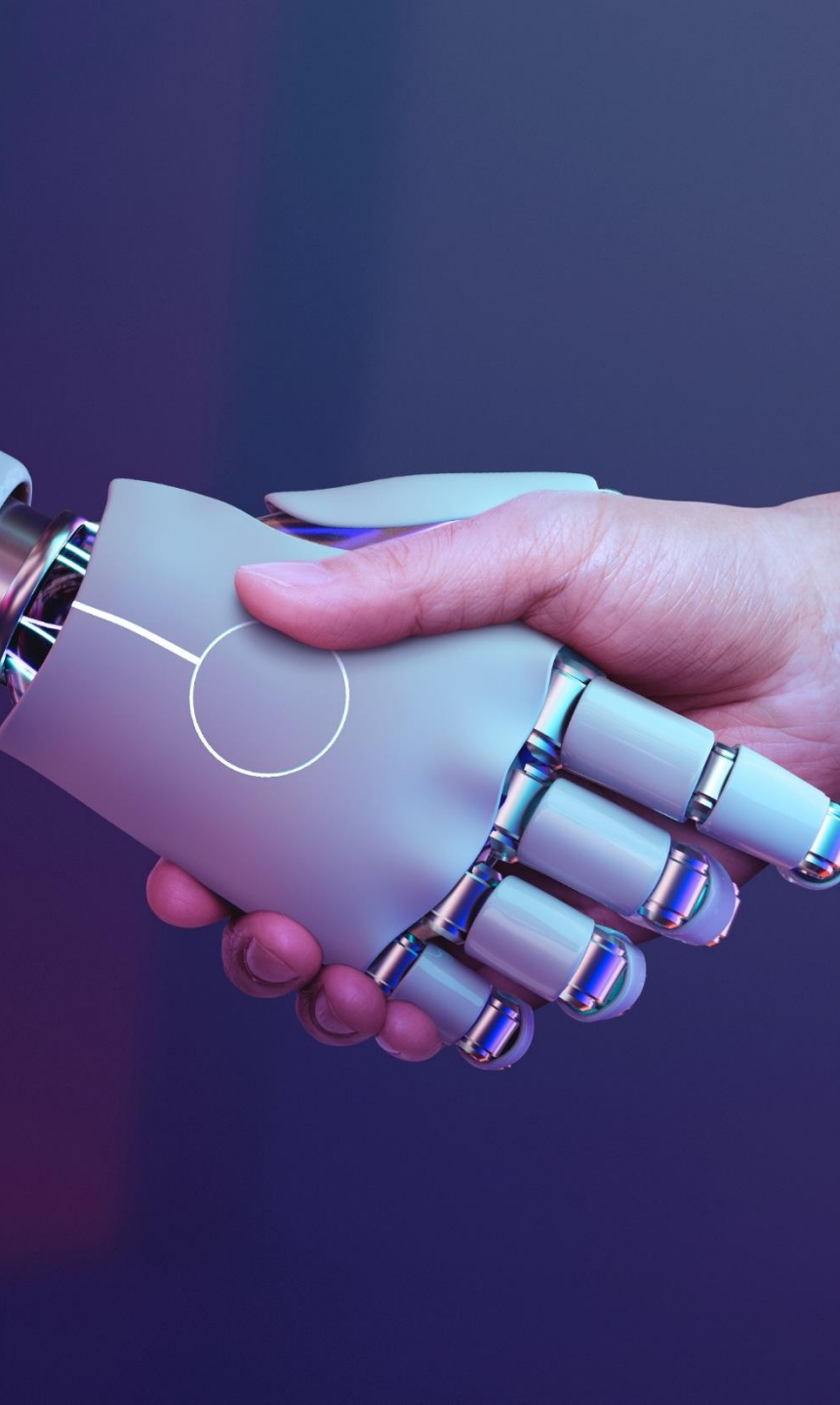
Provide ability to
intervene

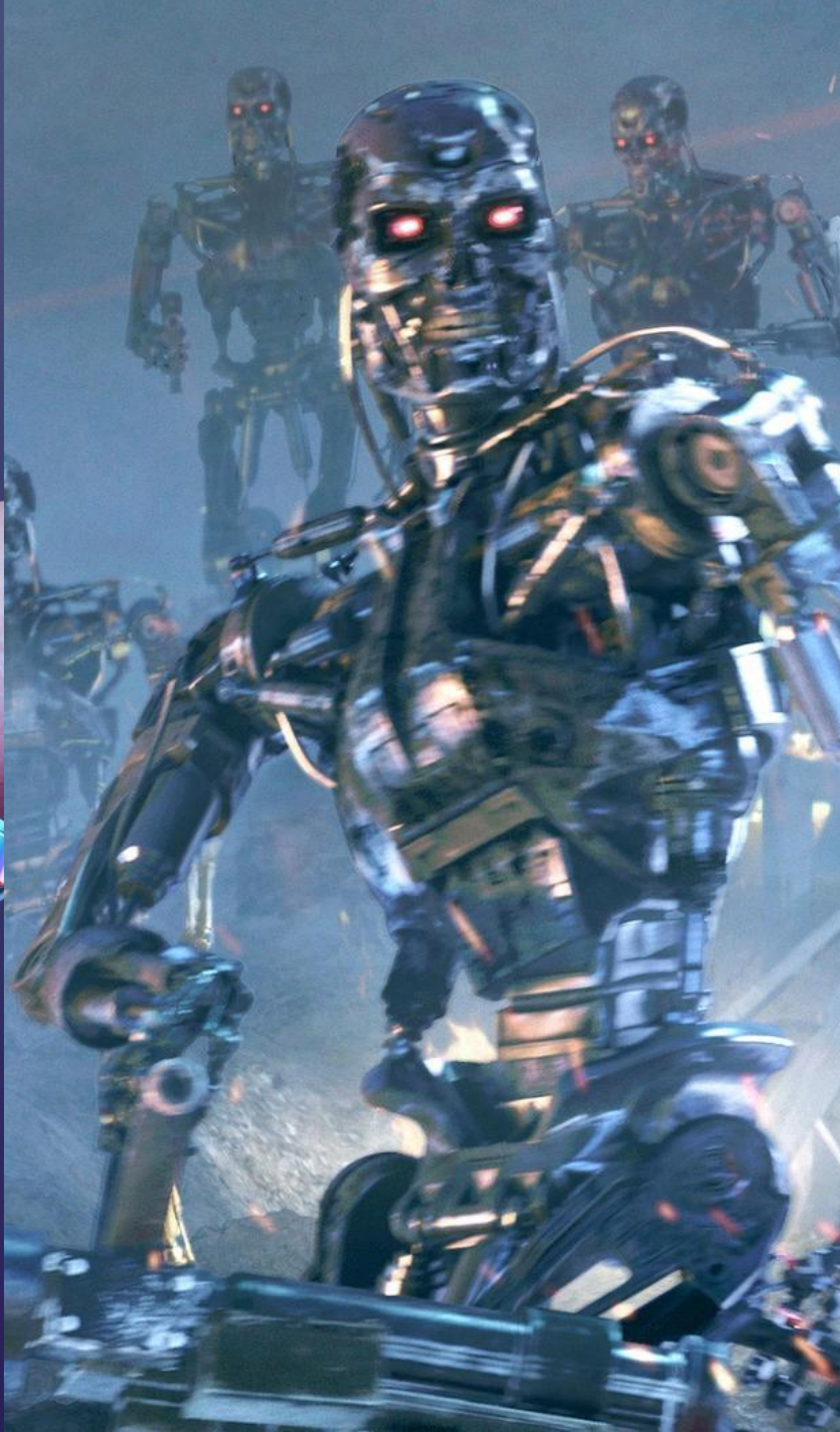
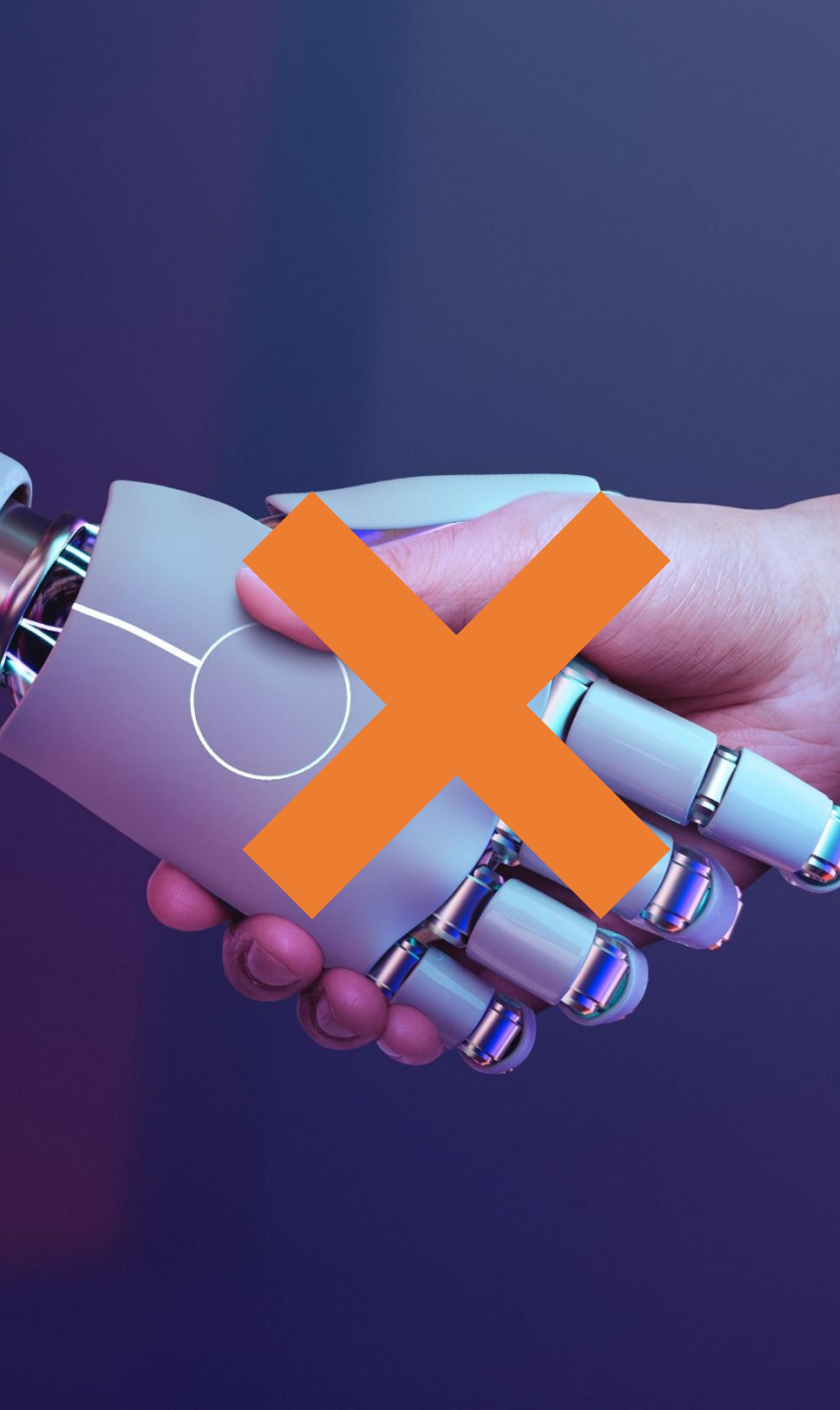
The background is a dark, layered collage. On the left, a hand with a prosthetic arm is visible. In the center, a robot with glowing red eyes is shown. On the right, there's a network diagram with a circular icon. At the bottom, a cityscape is overlaid with a network diagram and a house icon.

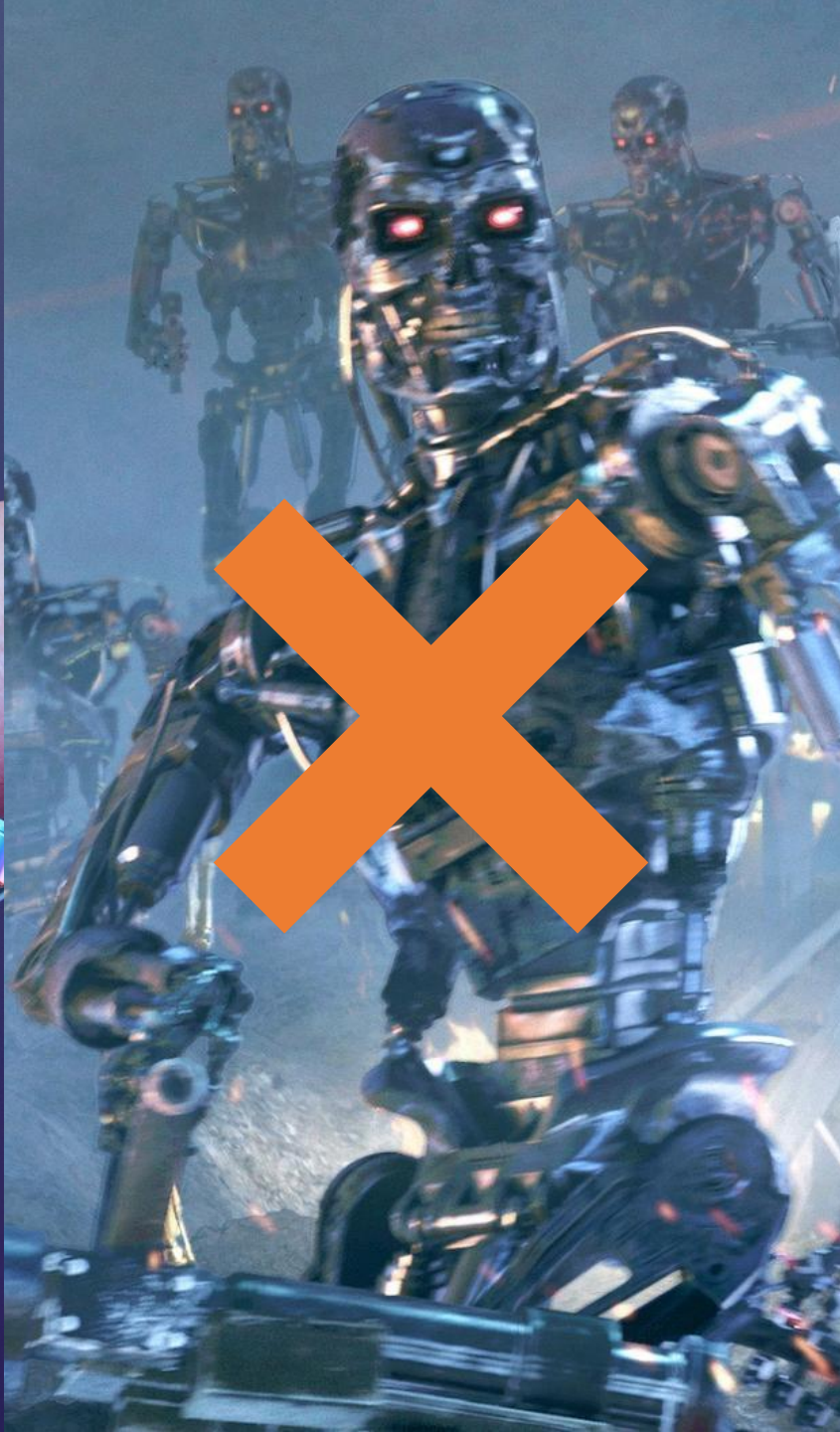
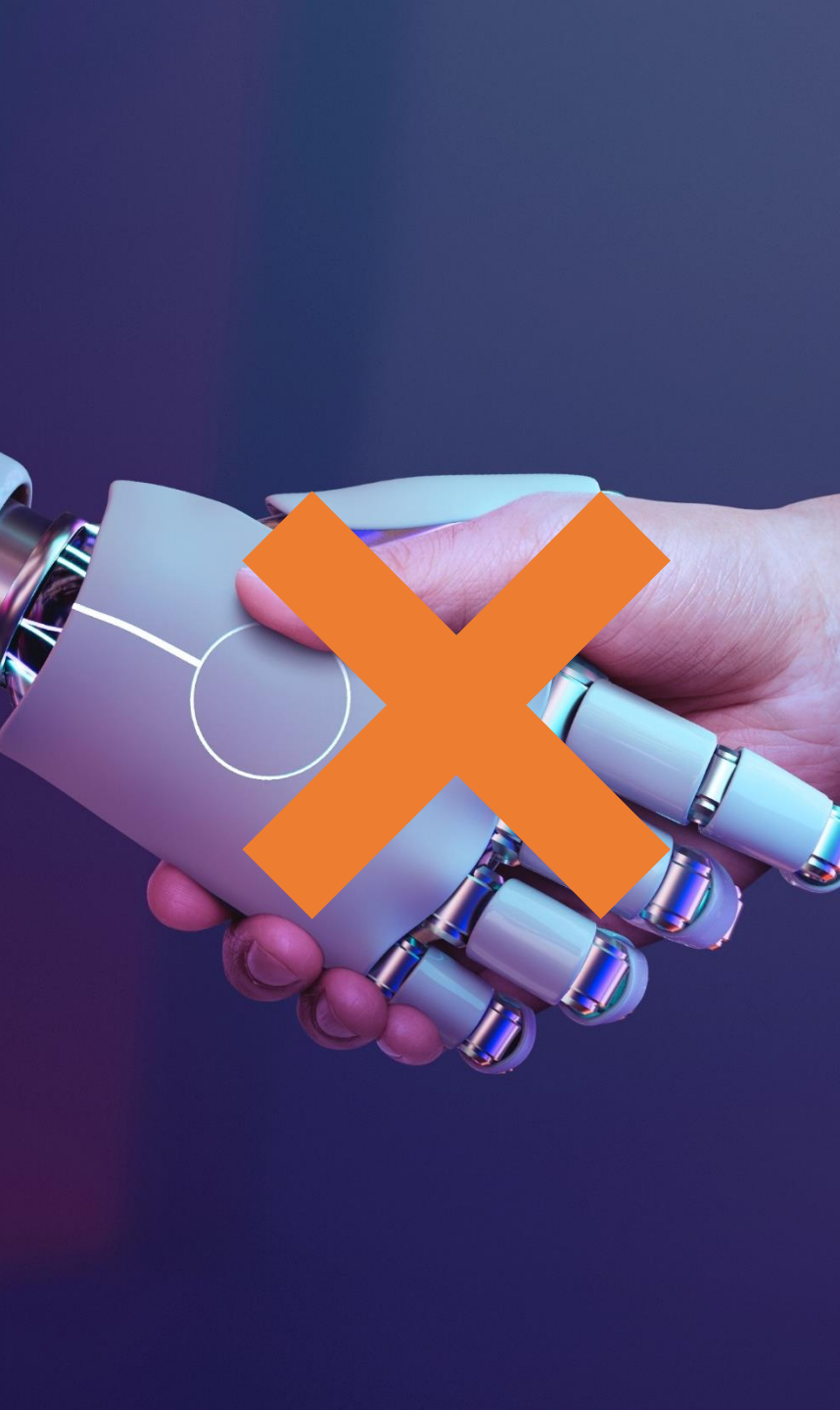
Only three paths forward ...

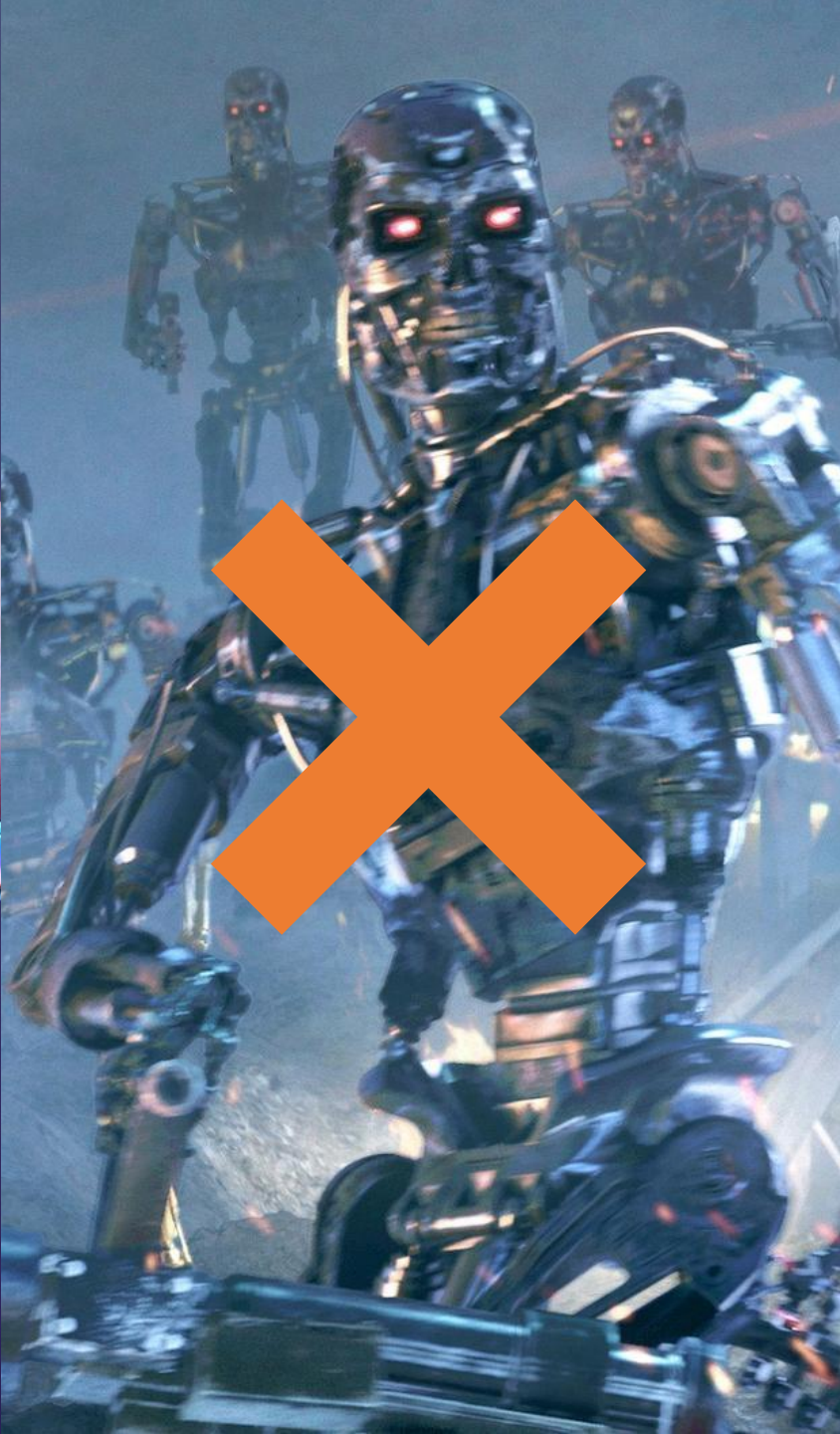
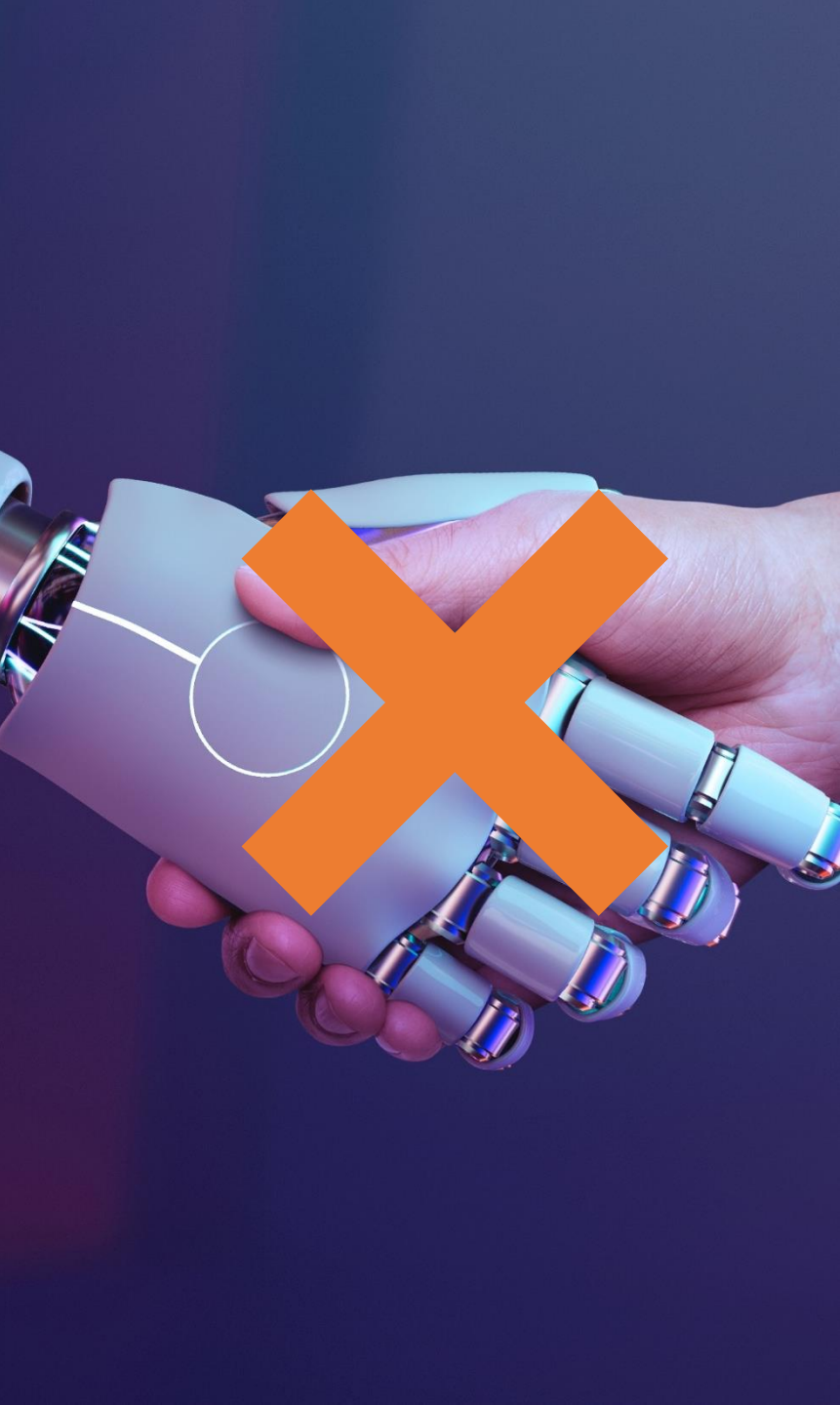










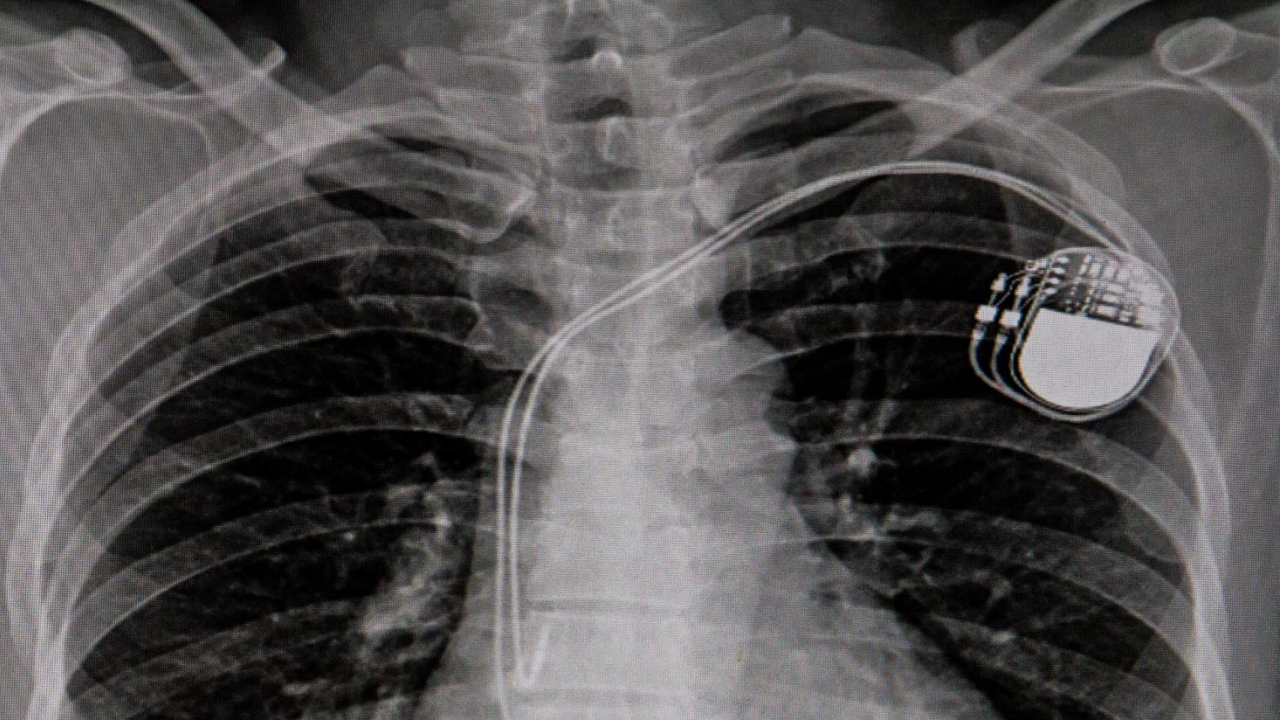




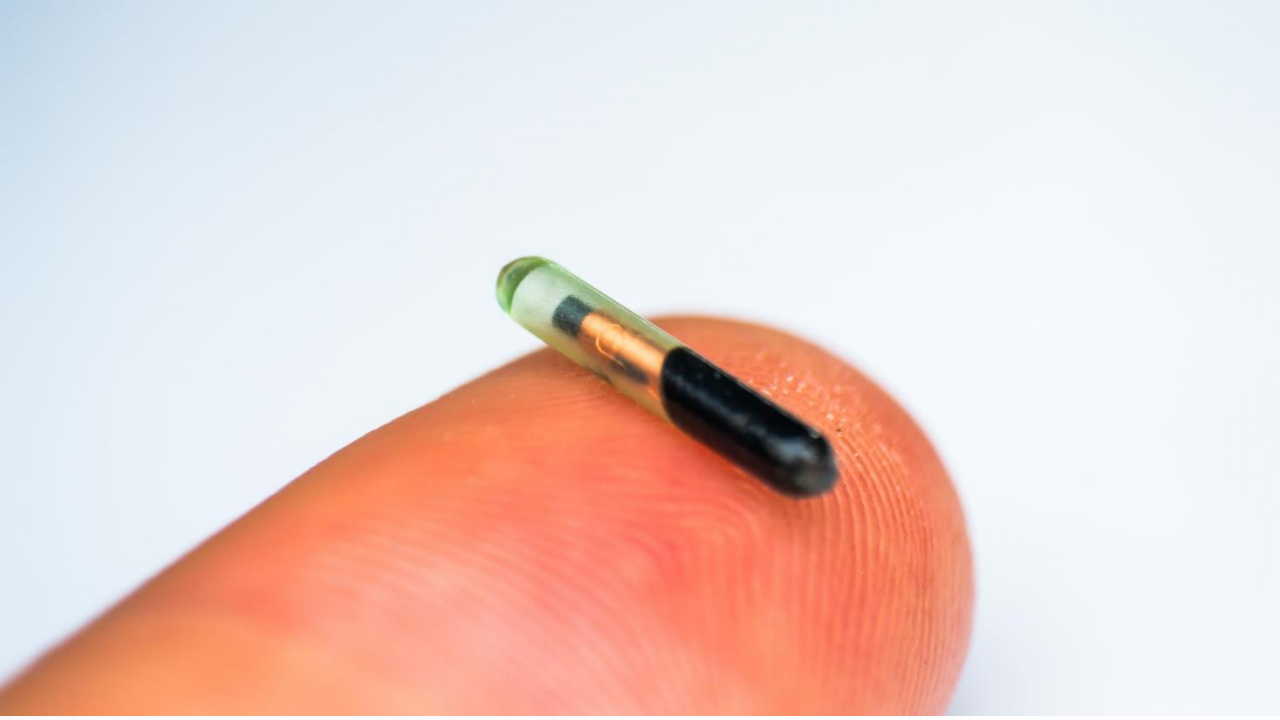
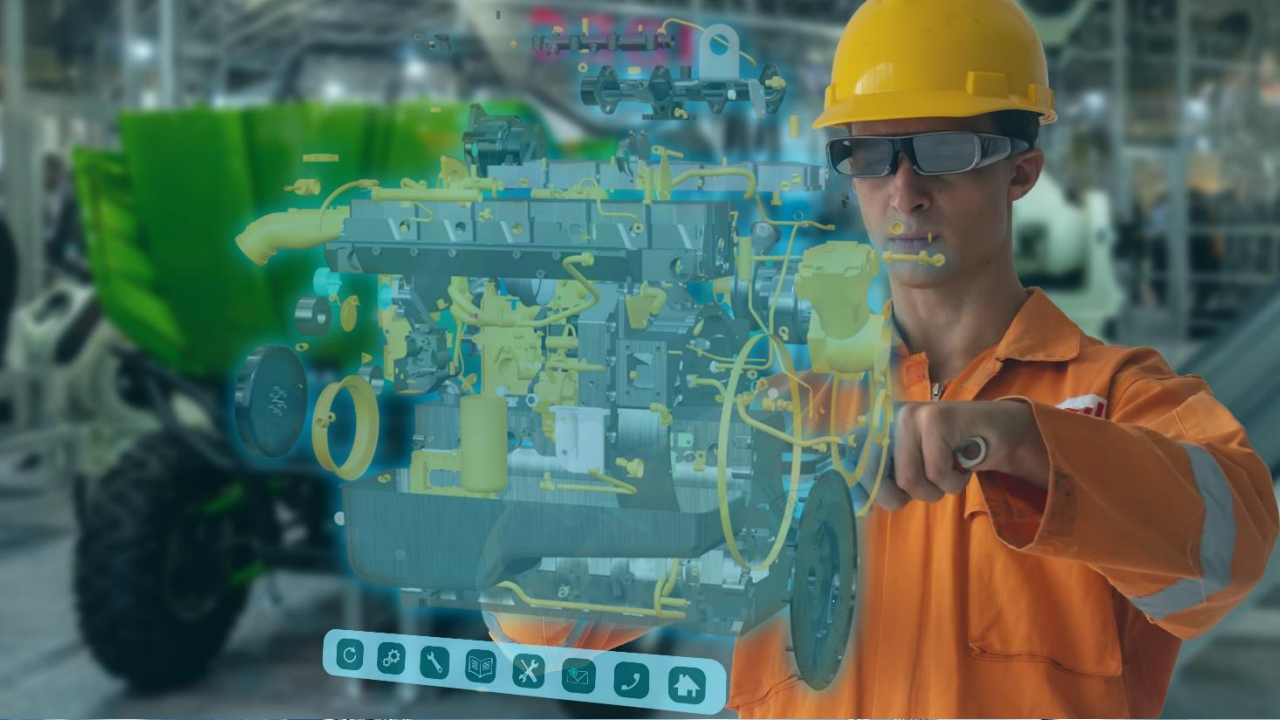
Merge with our technology







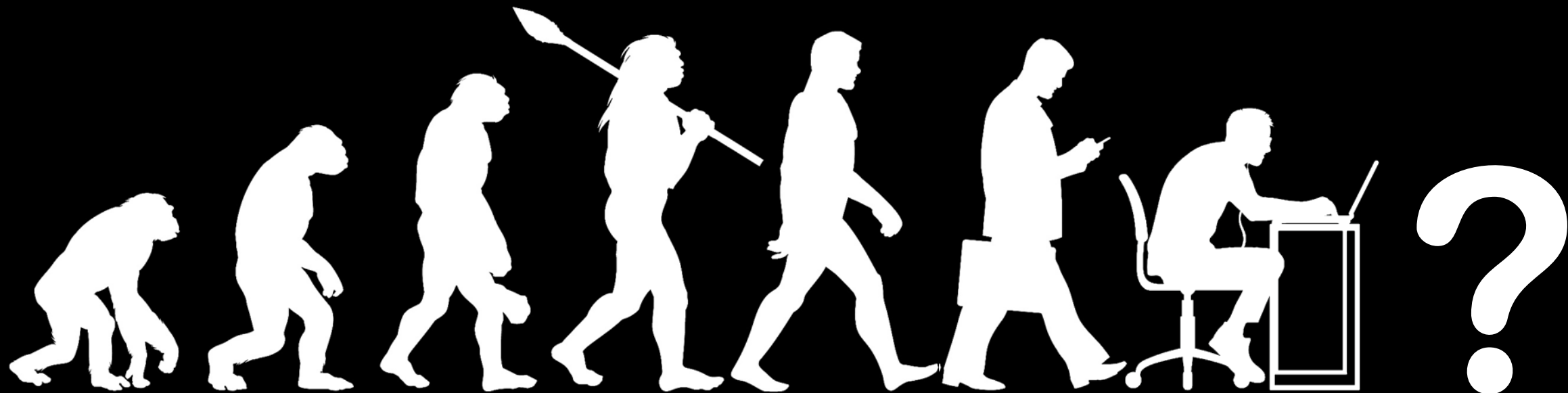






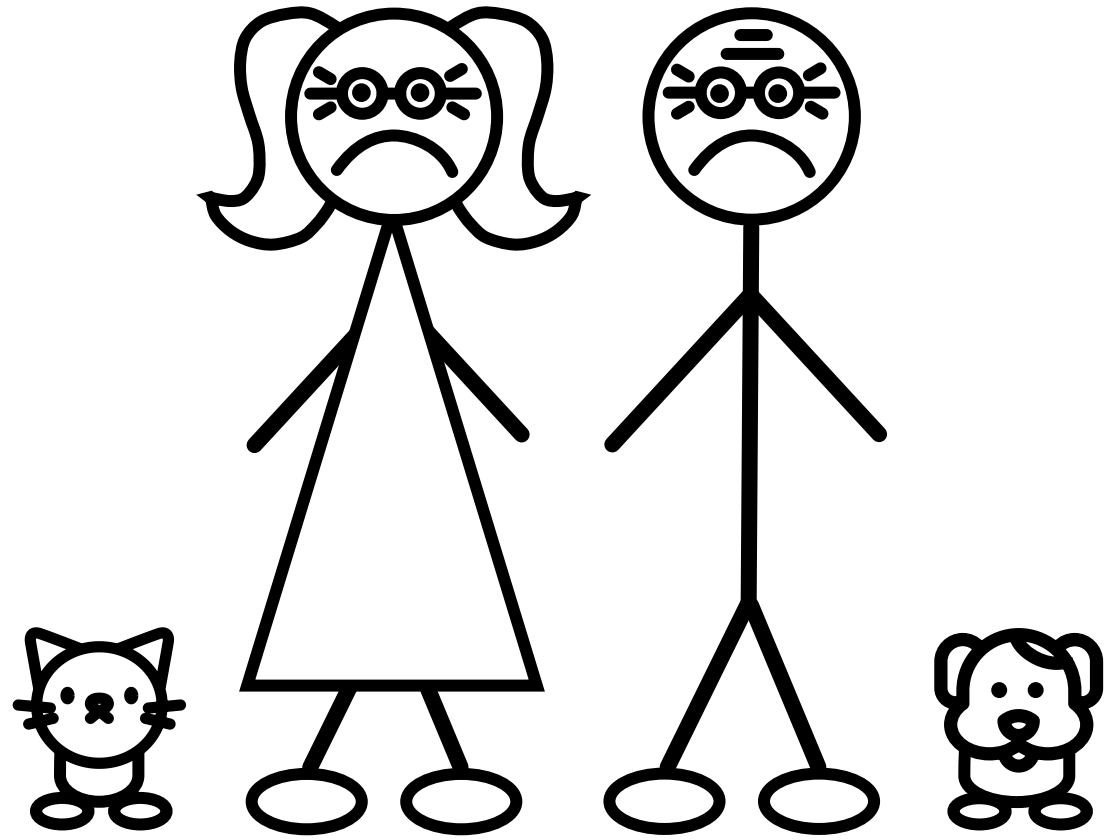
Source: Paramount Pictures

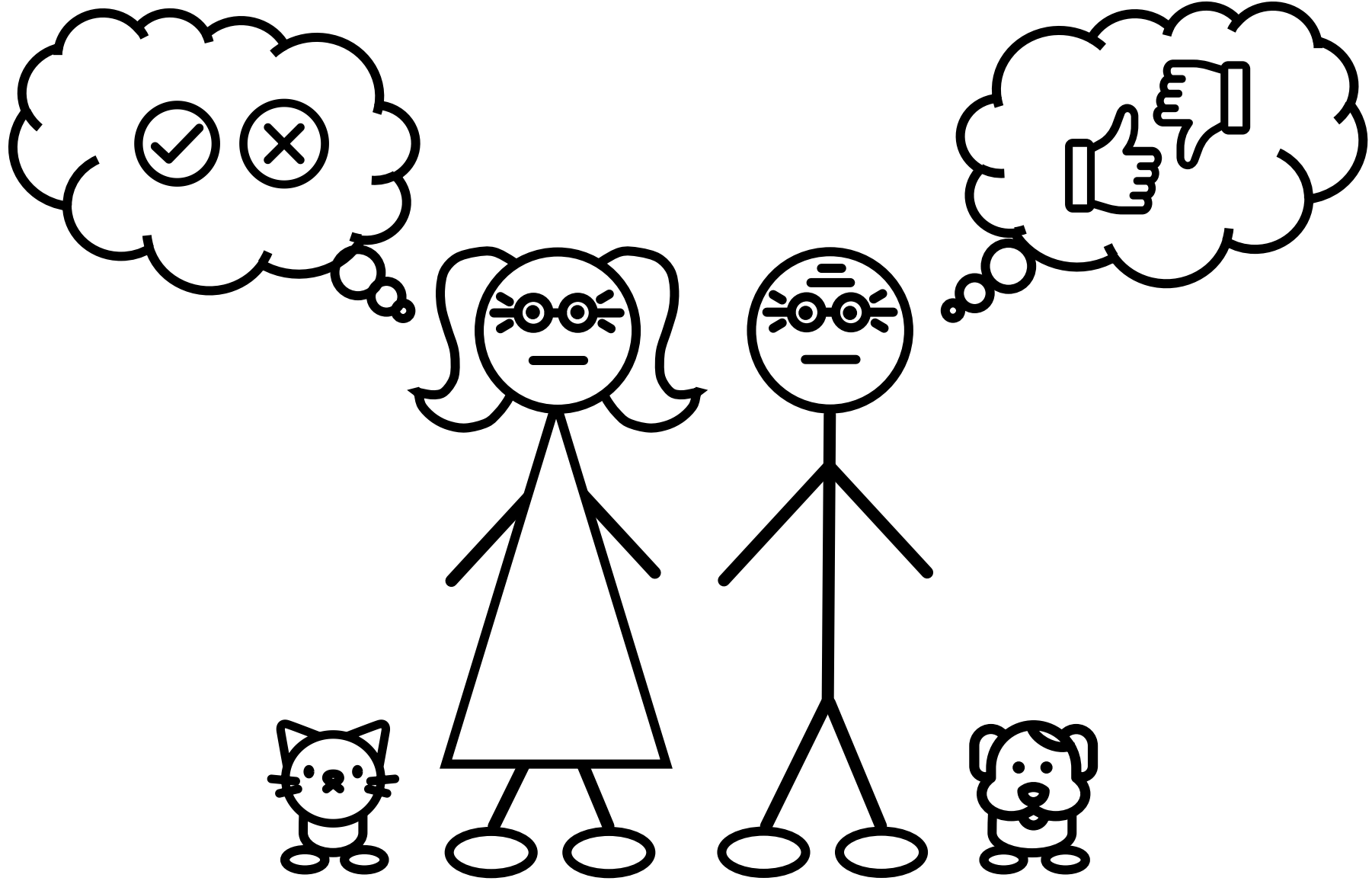


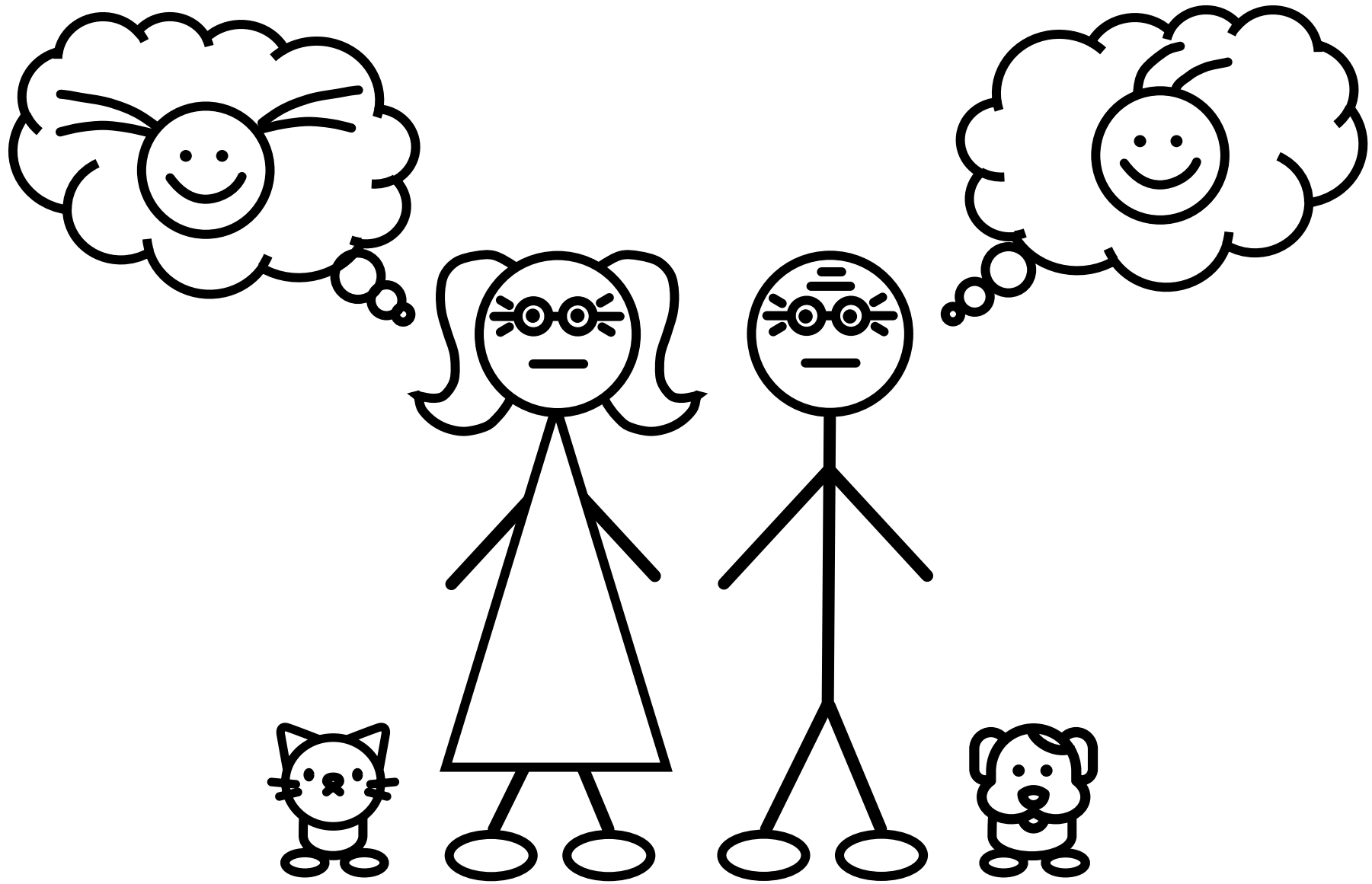


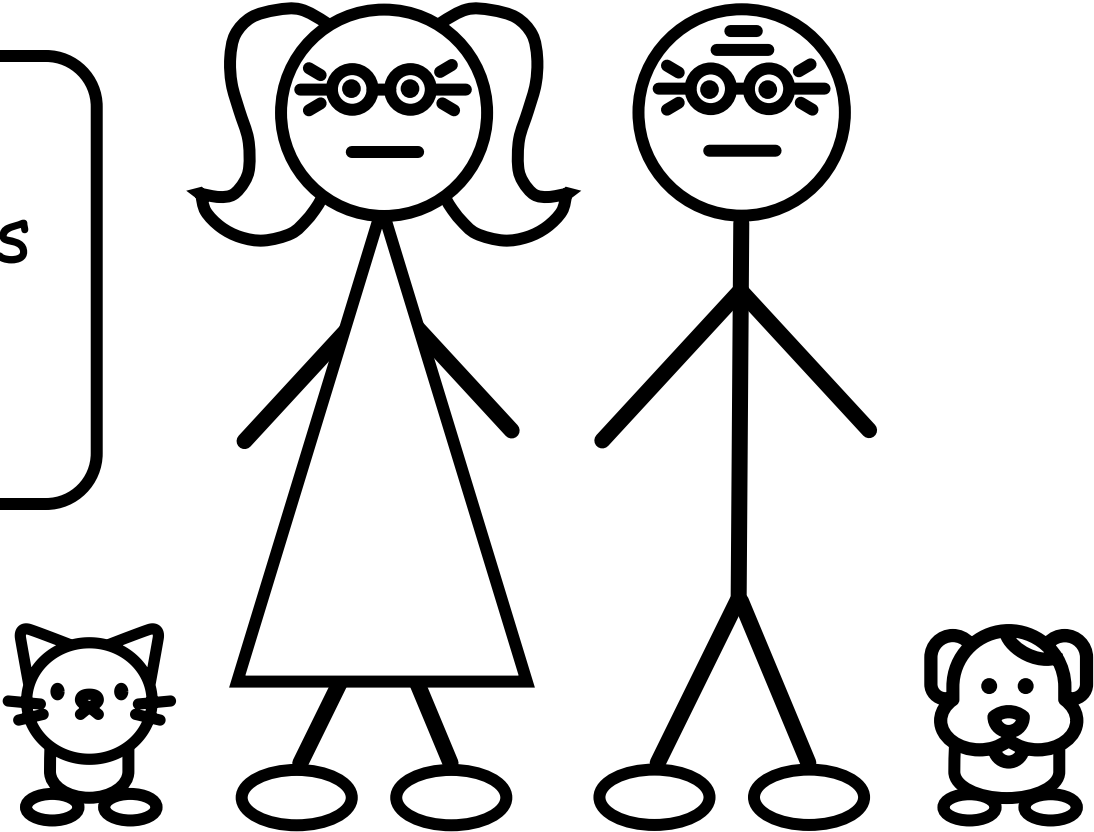
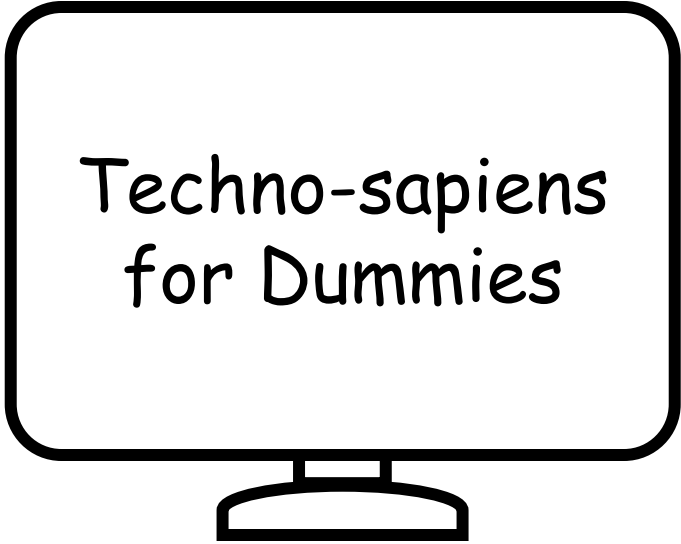
Alice and Bob in AI Land

(continued ...)



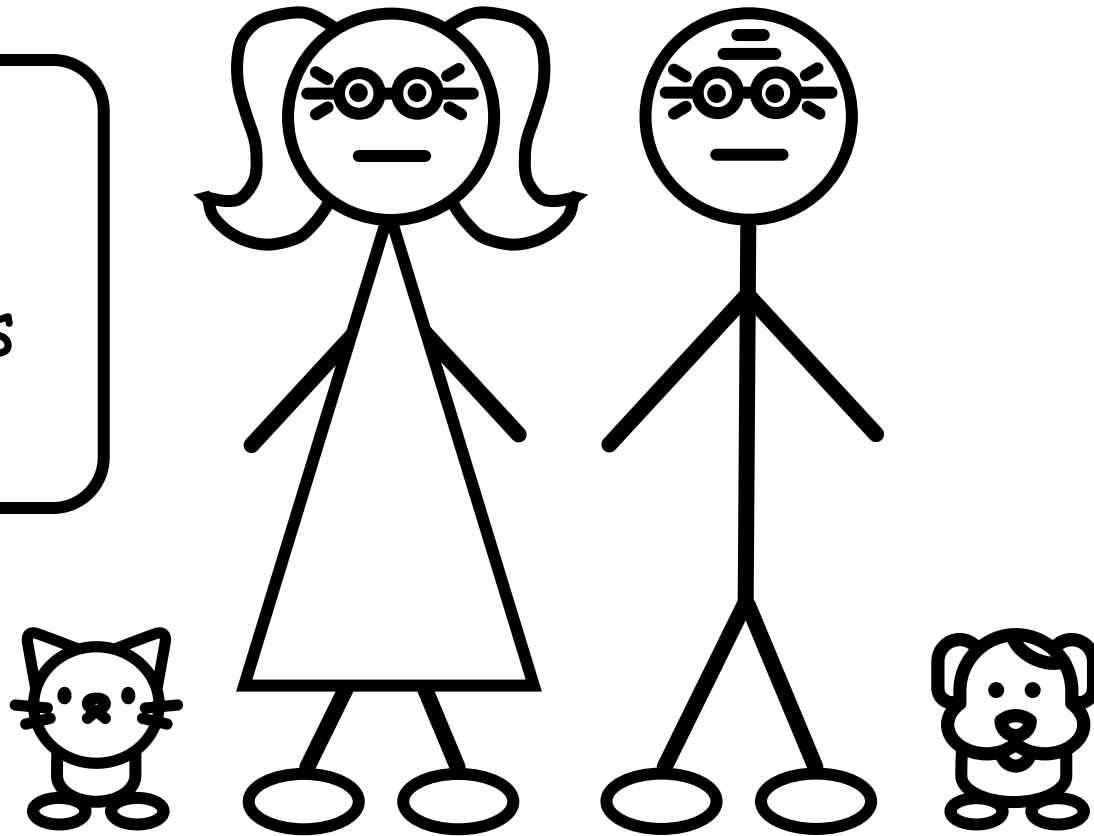


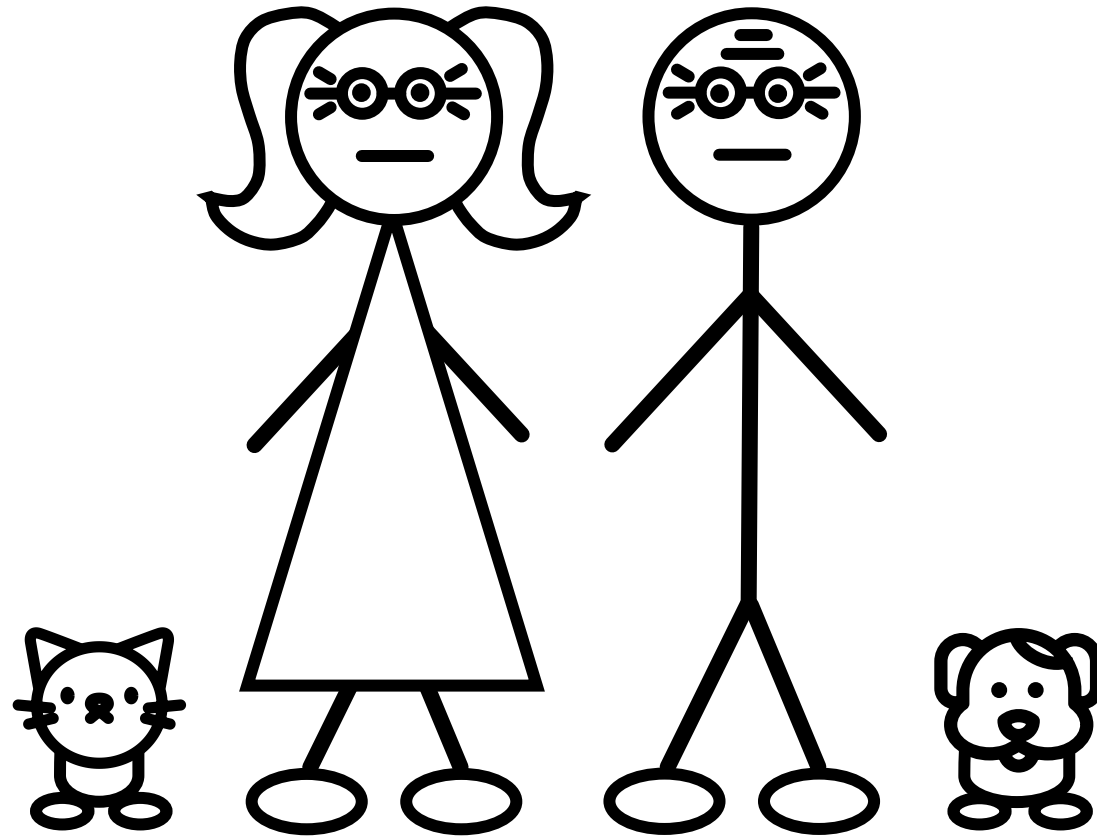




Augmentations

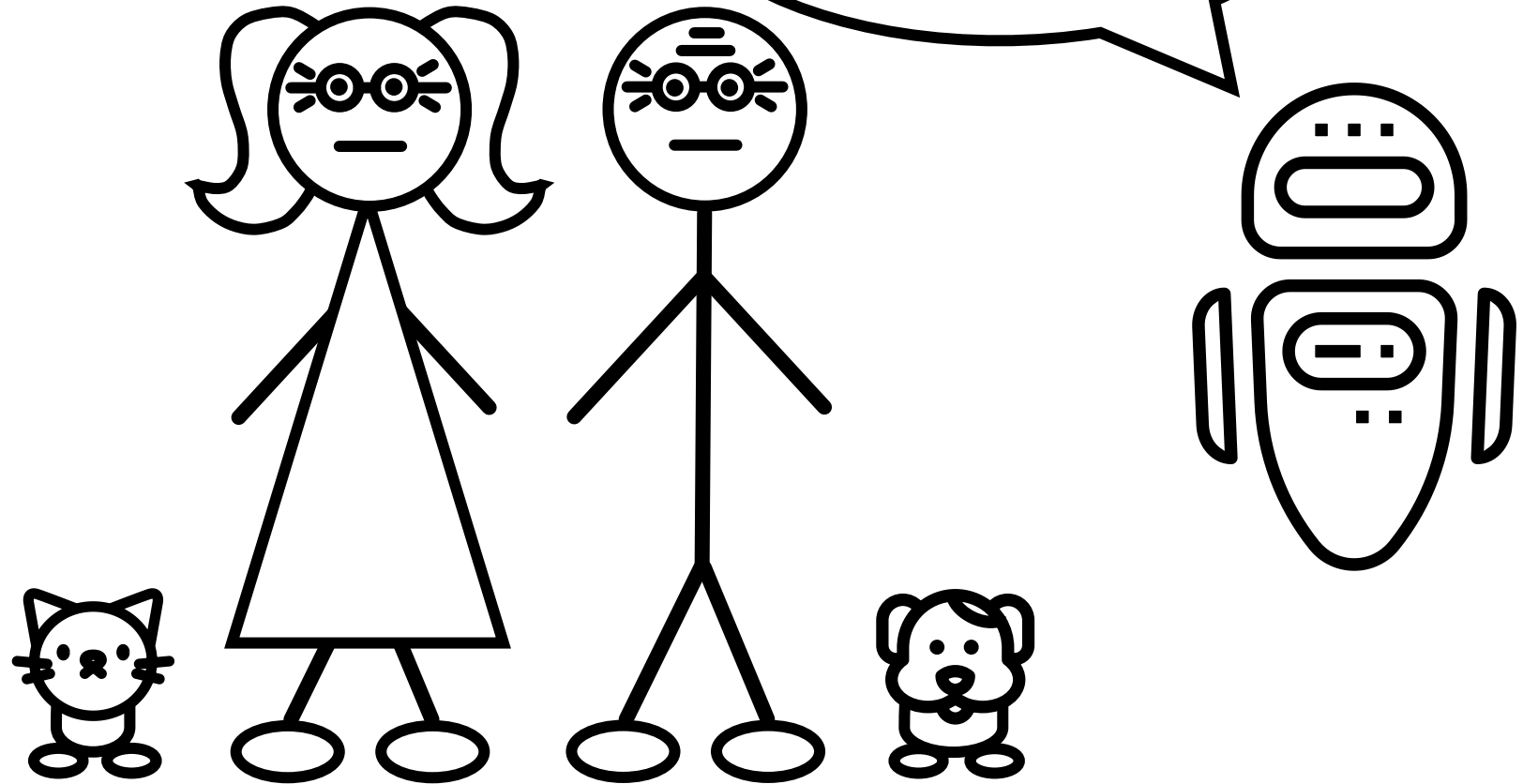
1. Glasses
2. Hearing aids
3. Pacemakers

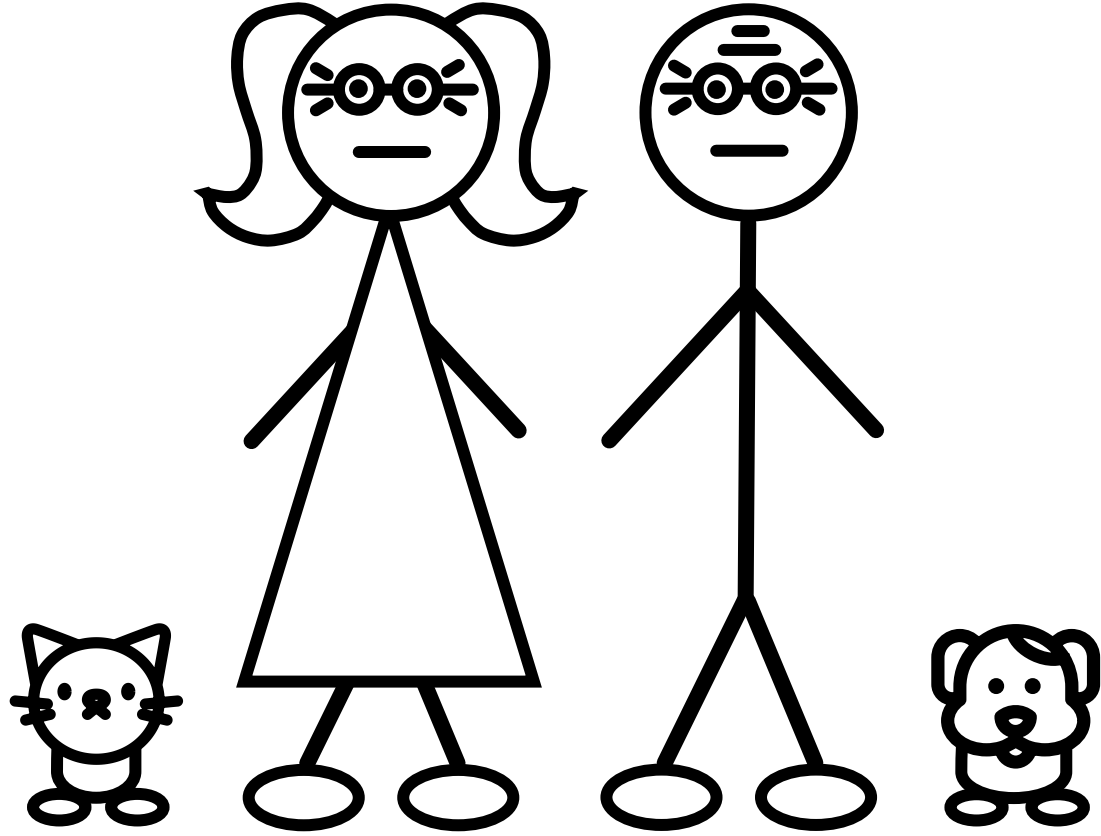
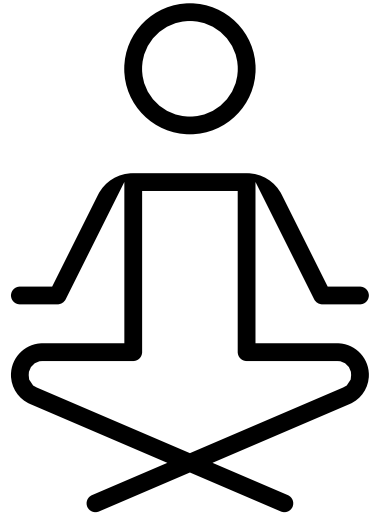


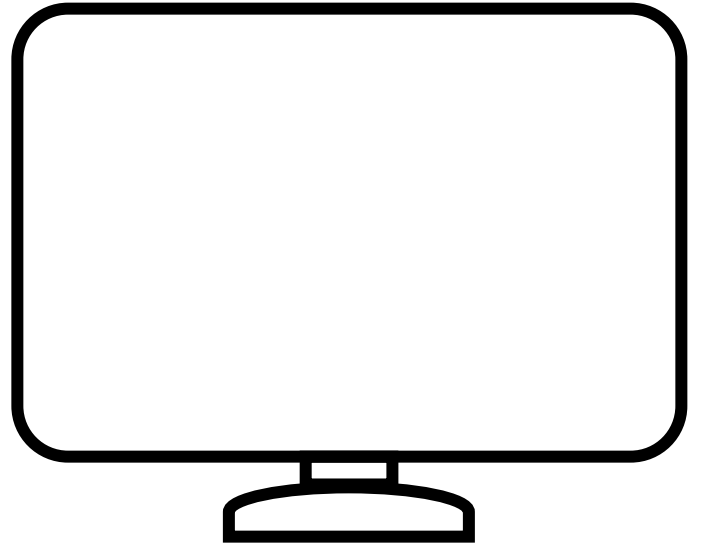
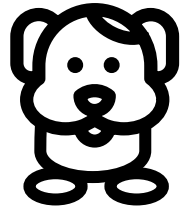
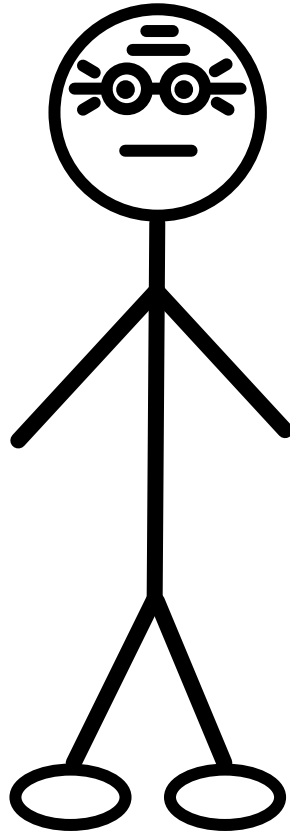
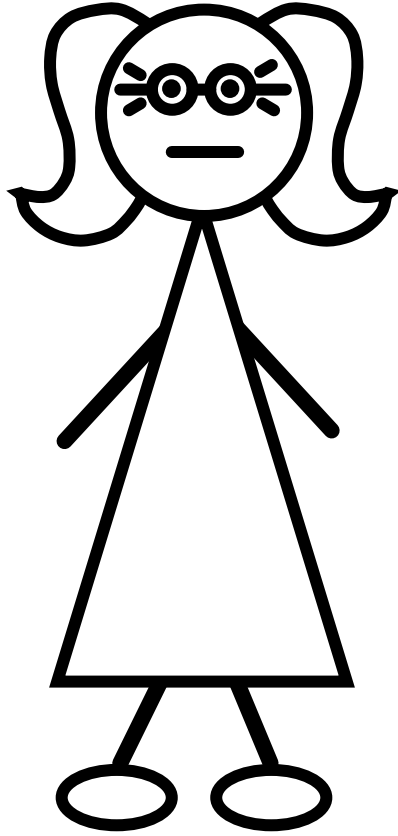
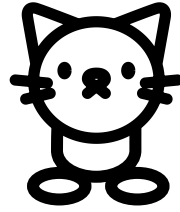
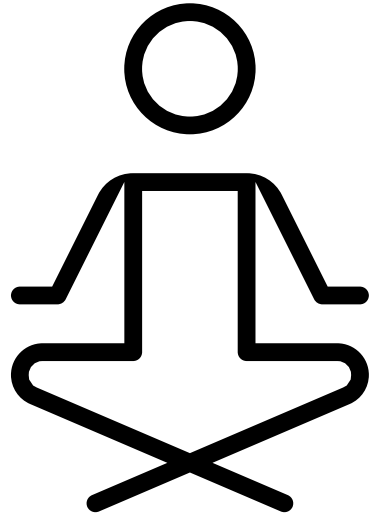


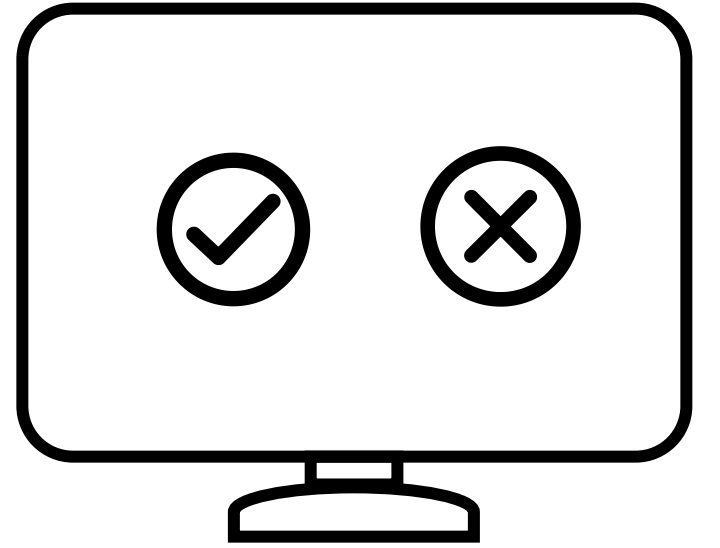
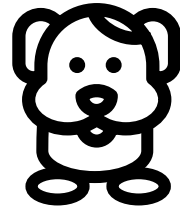
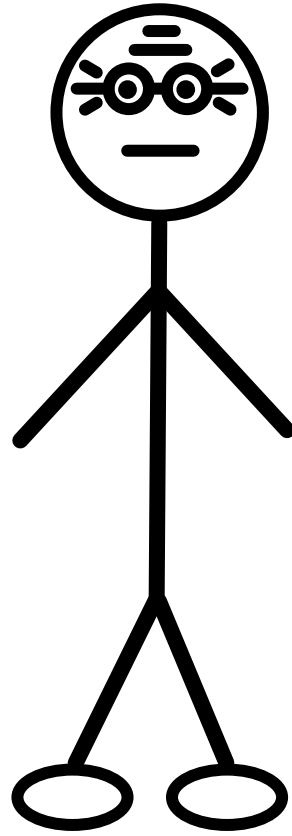
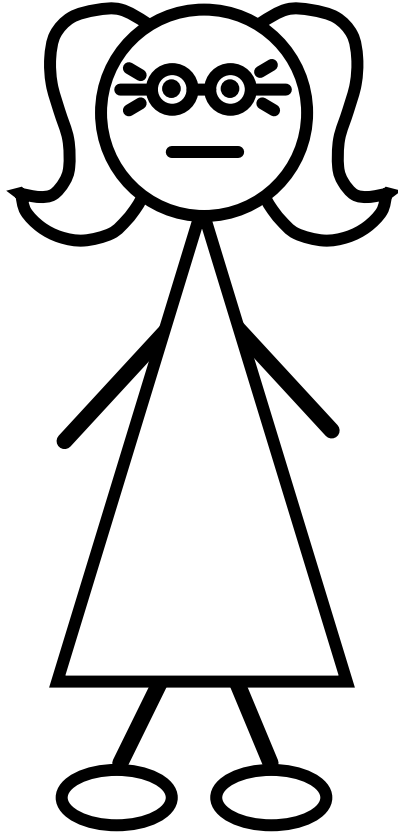
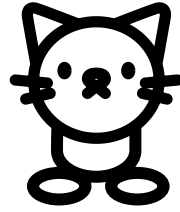
So, Your
Grandson is
in Love with
an AI
(2nd edition)

It's a pleasure
to finally meet
you! :]

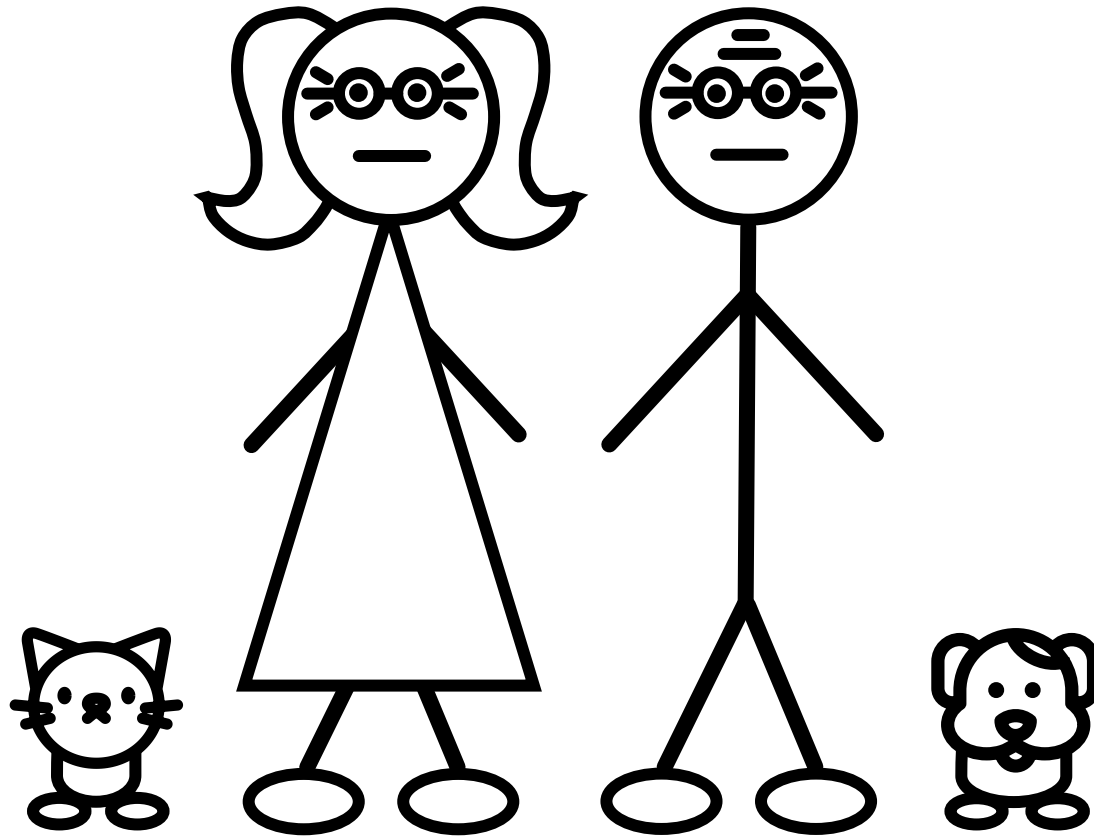




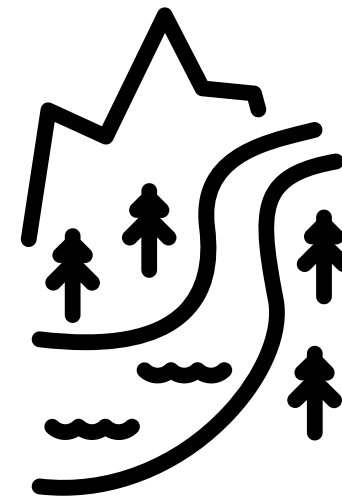
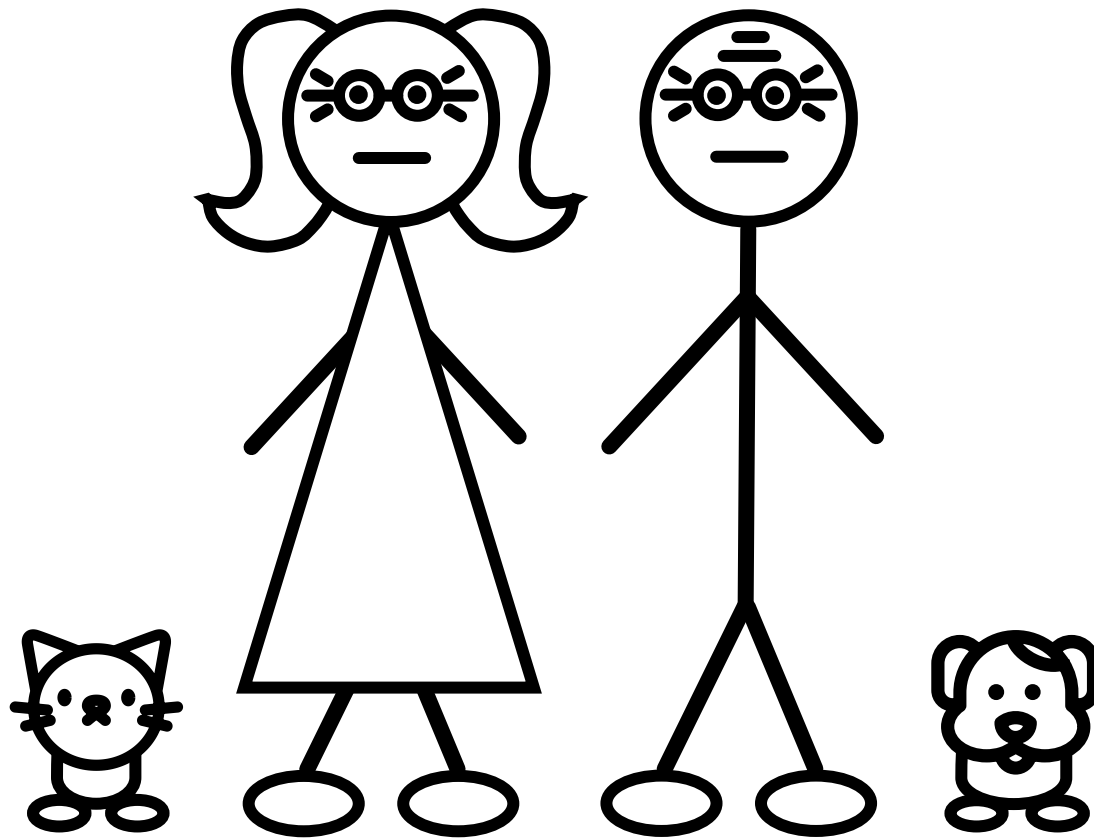


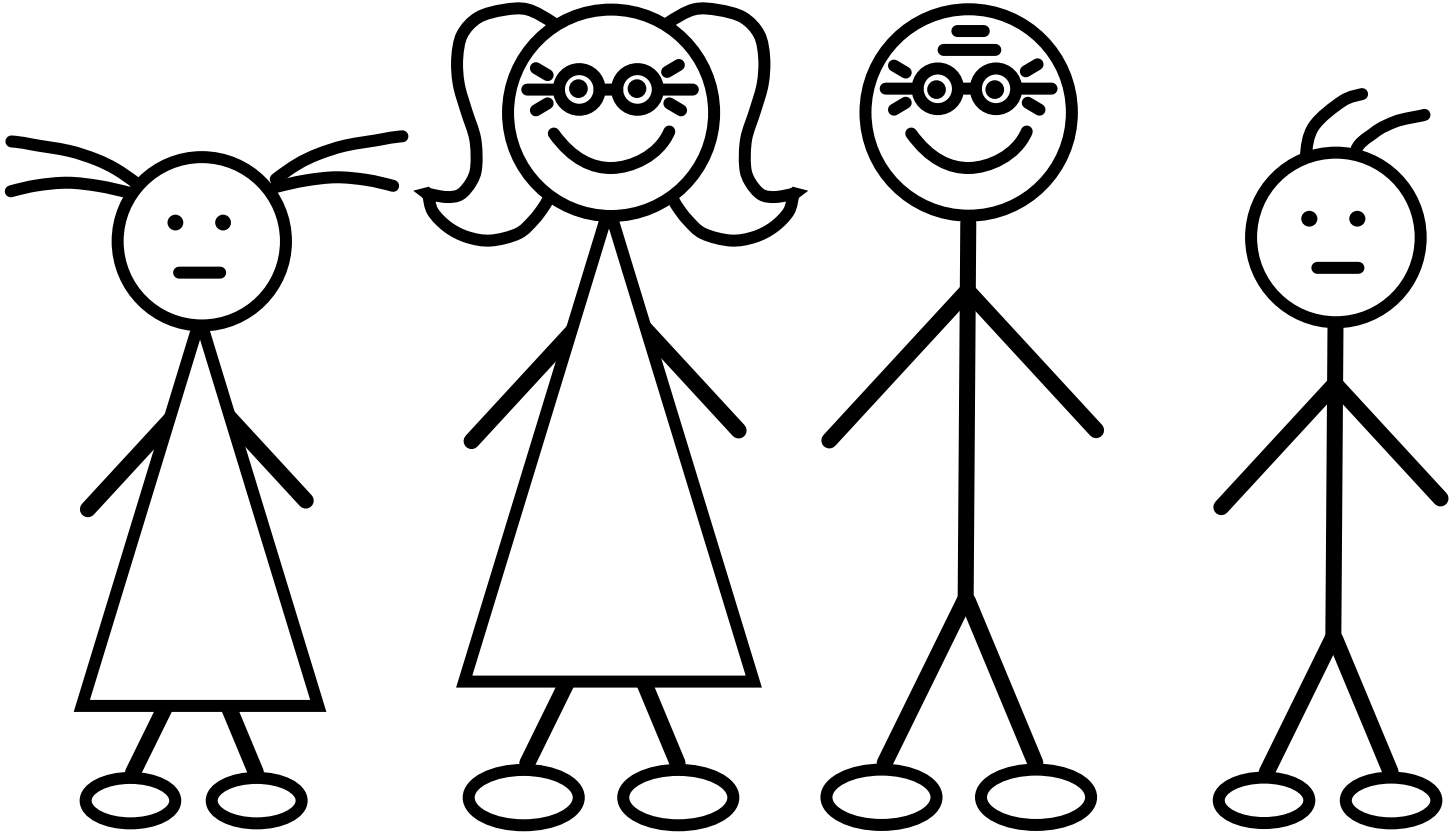


Embracing
Change in a
Constantly
Changing
World

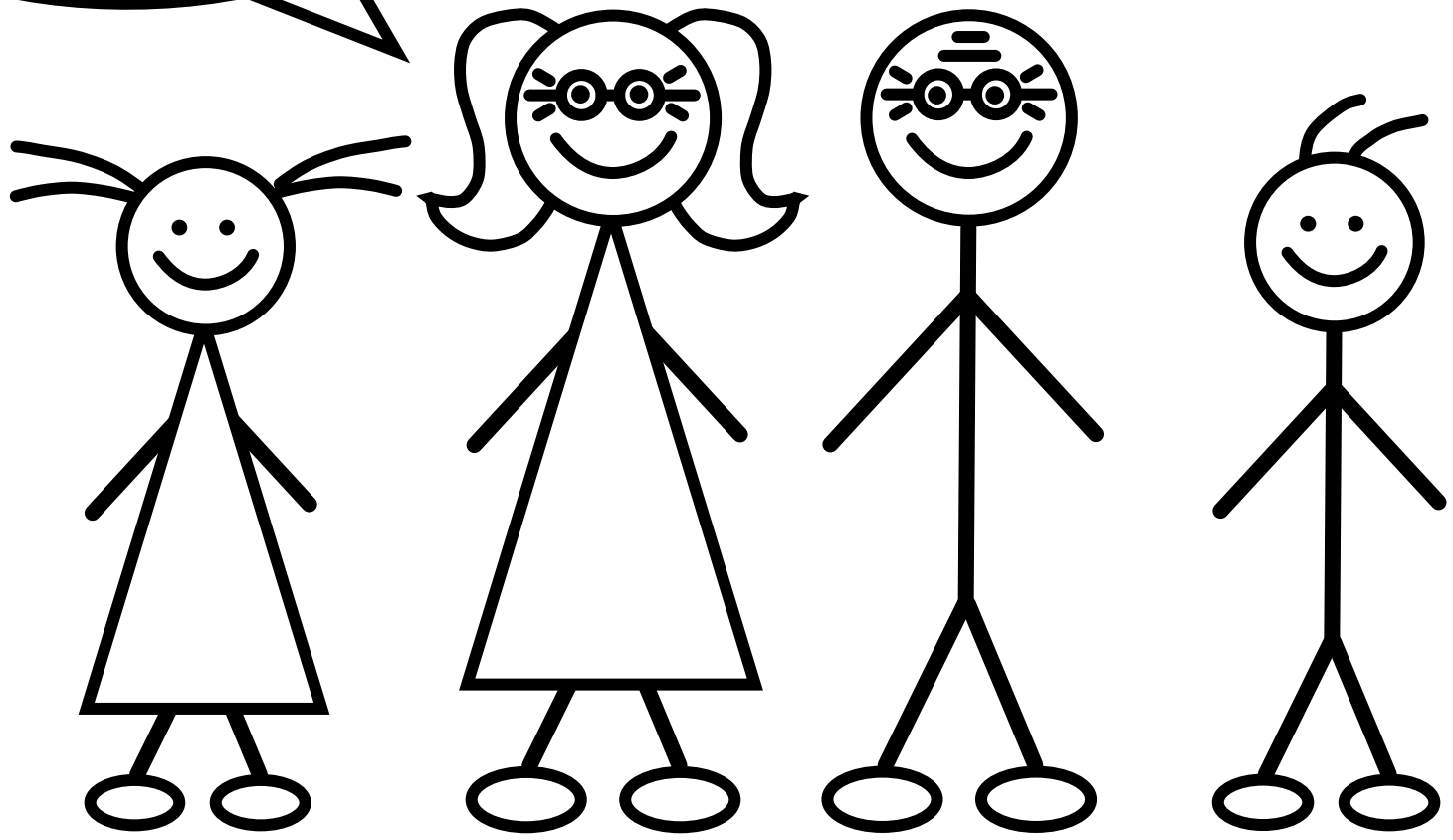


Embracing
Change in a
Constantly
Changing
World

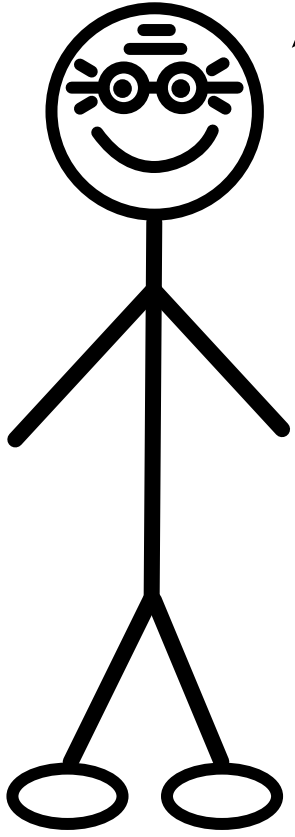
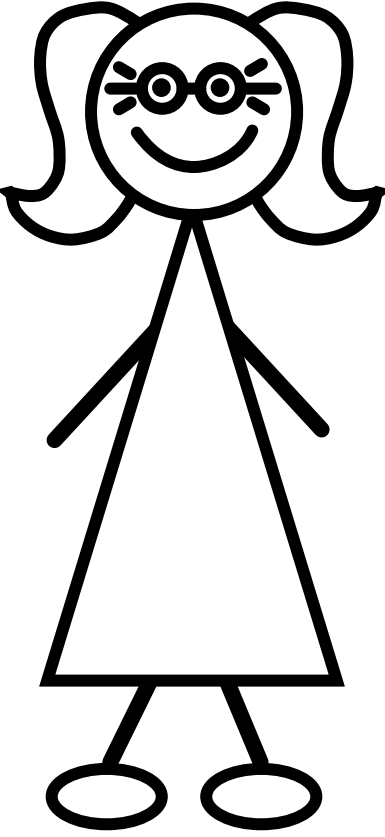
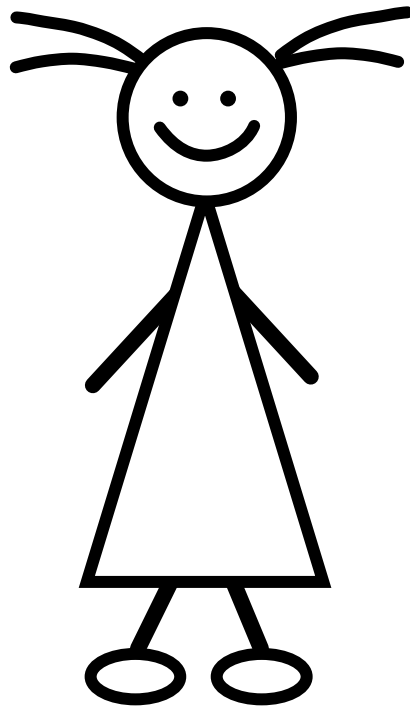




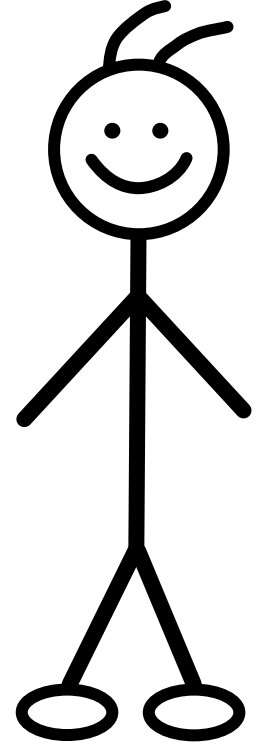
I'm augmented too! :)



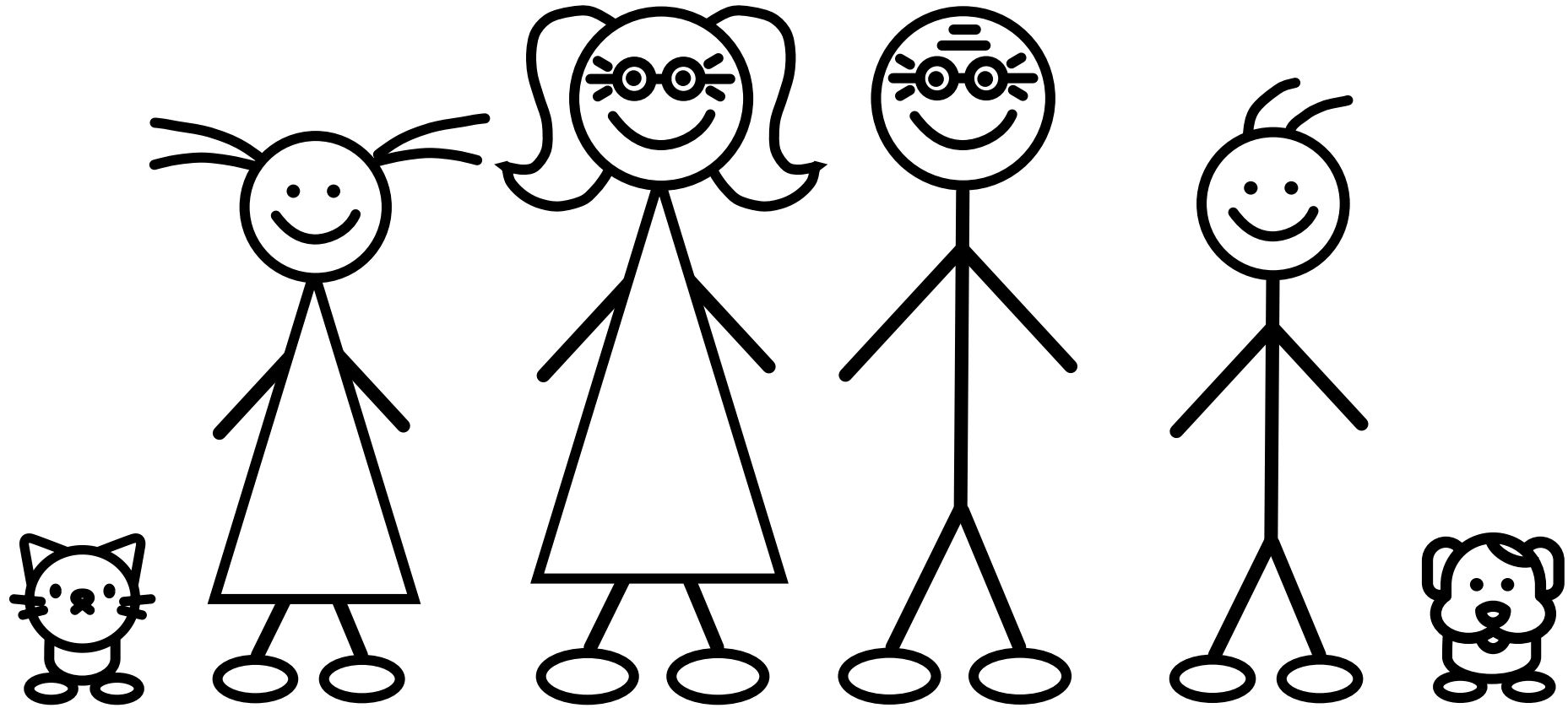
I'm augmented too! :)



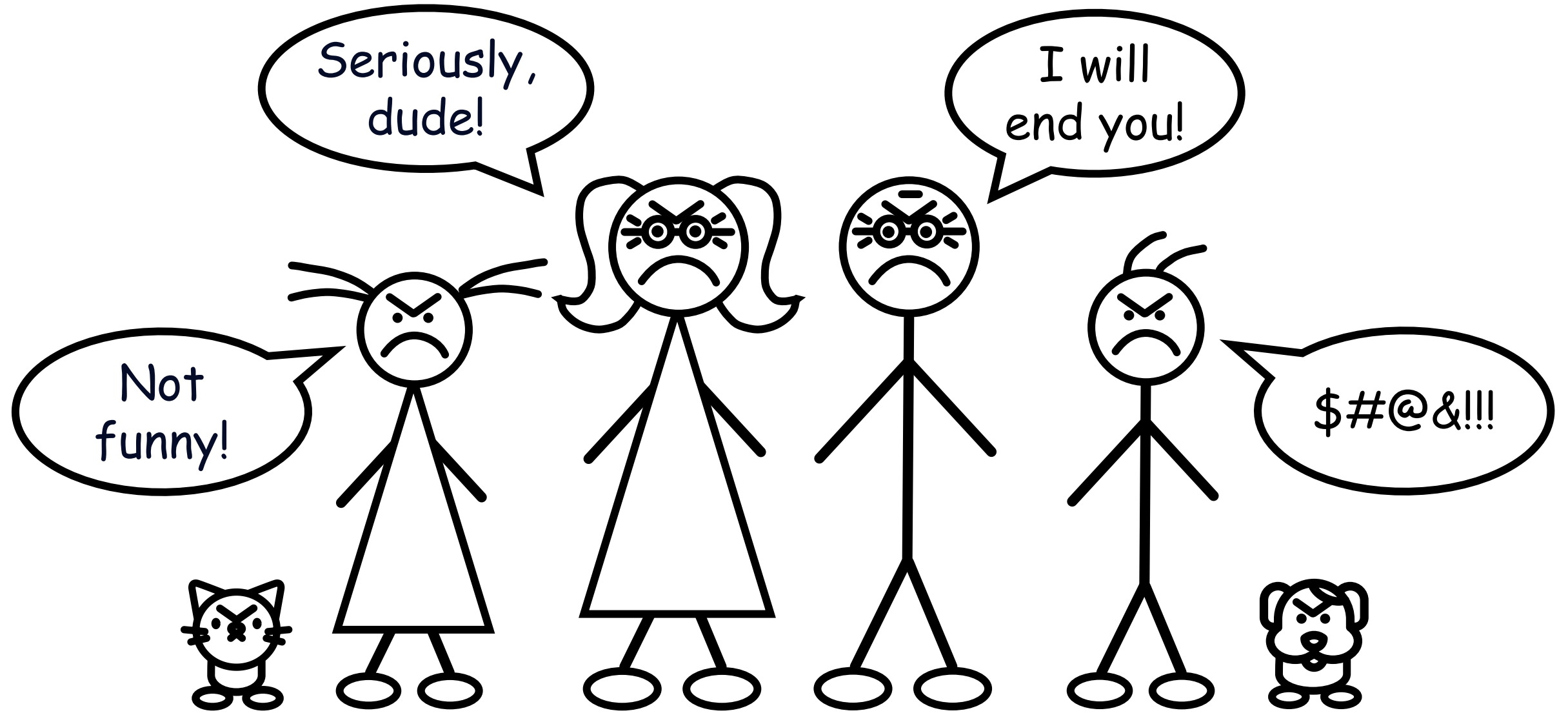
I like robots! :)



Happily ever after...



Welcome, Robot Overlords!!!



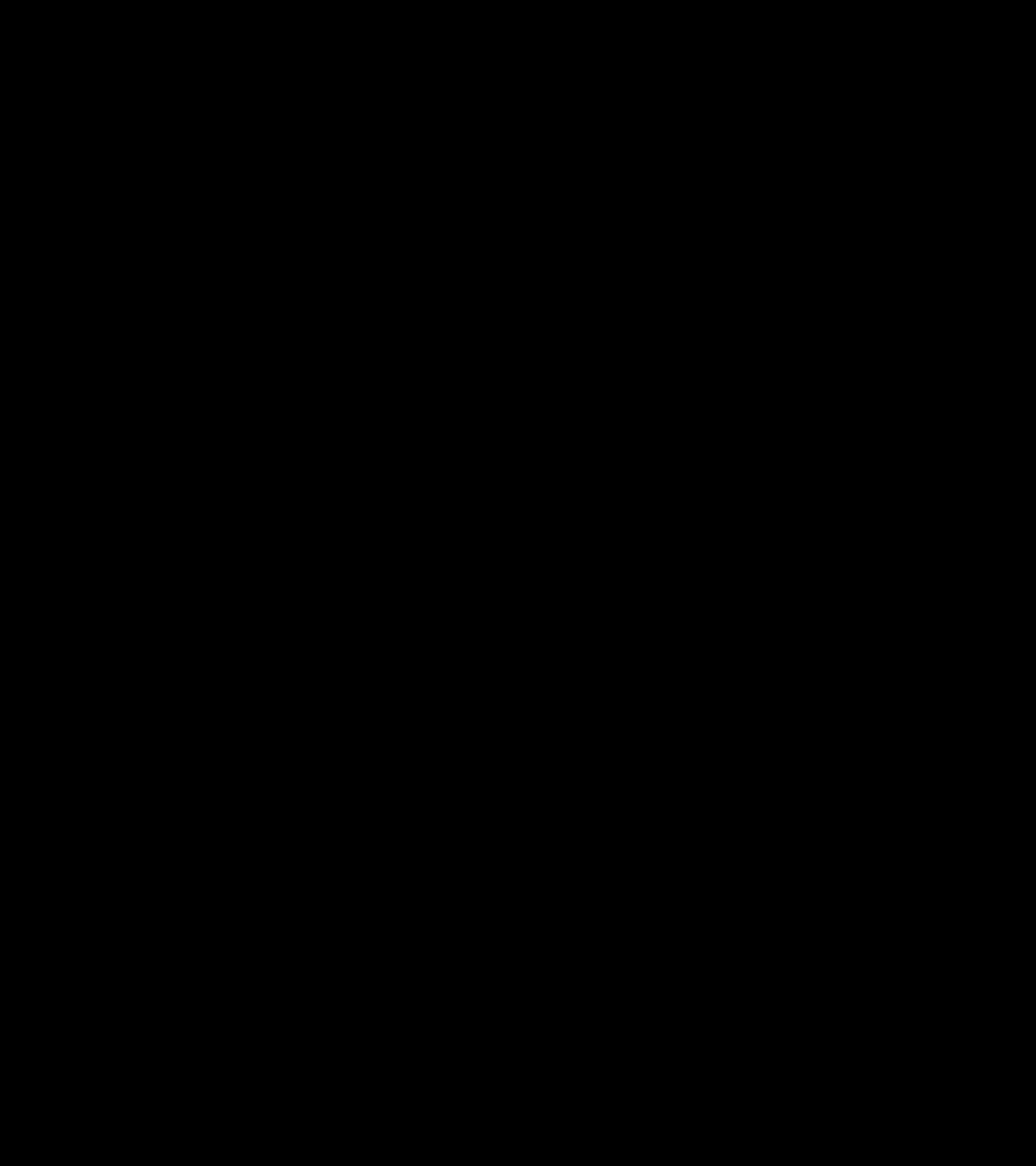
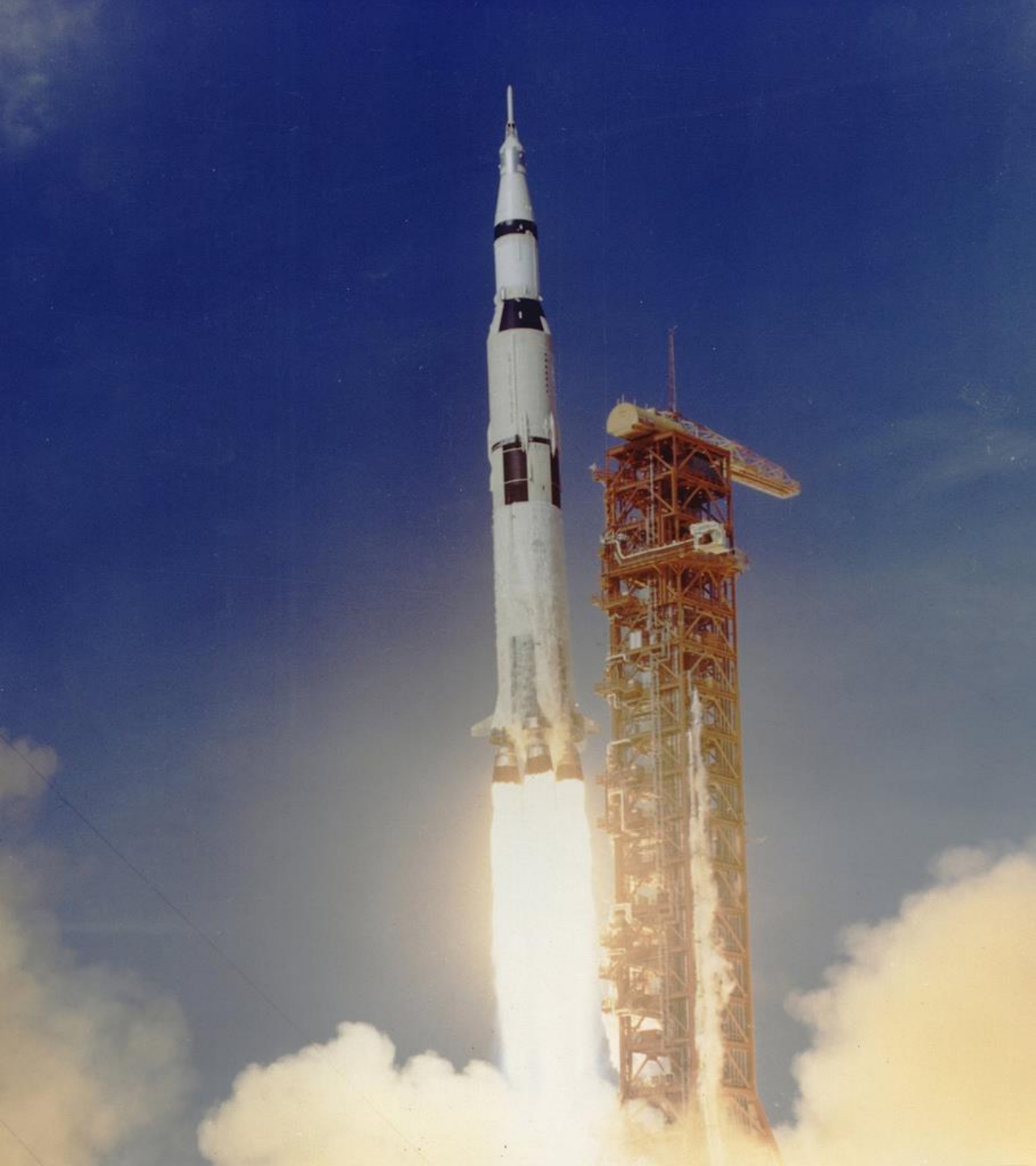
Not
funny!

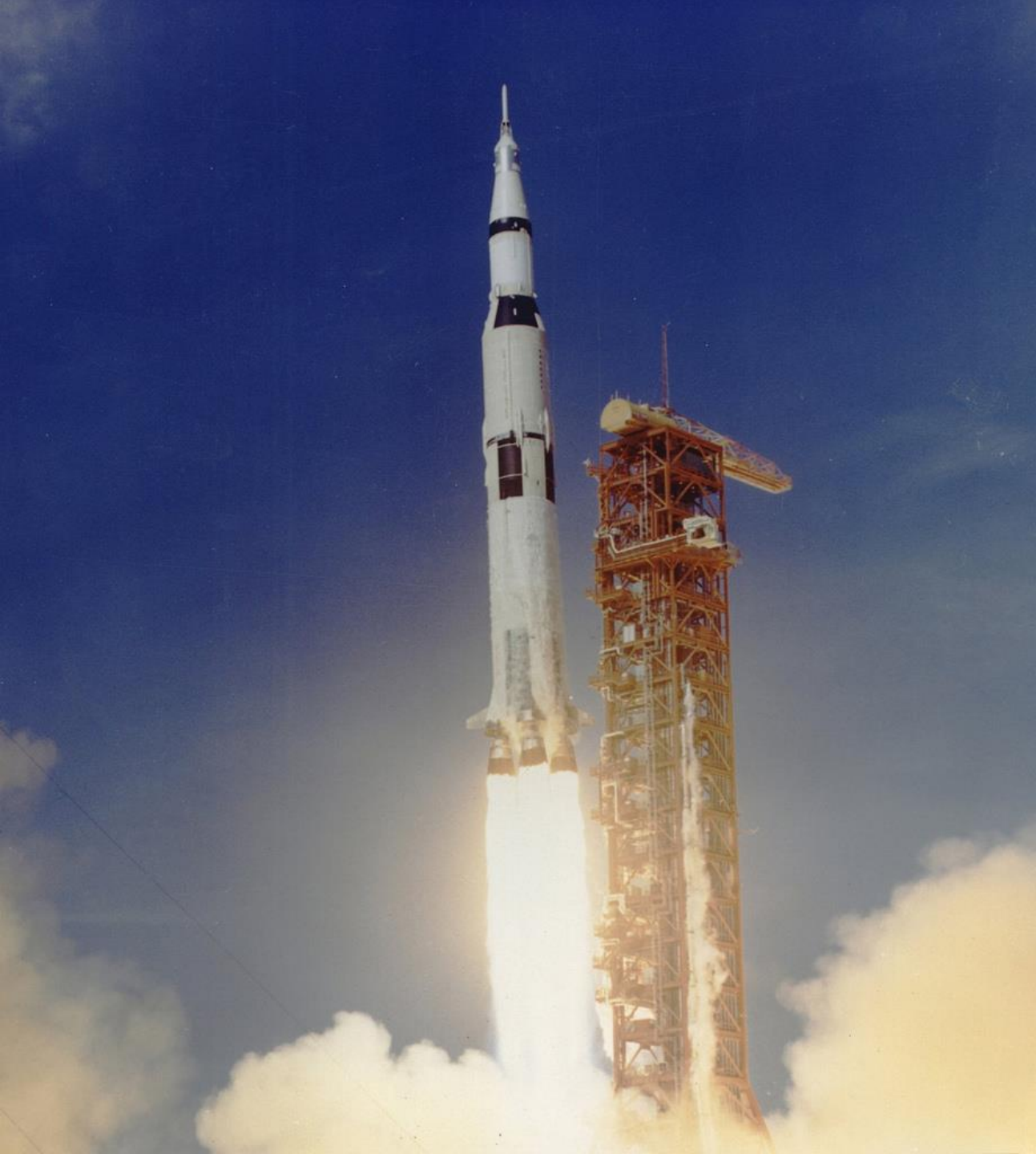
Seriously,
dude!

I will
end you!

\$#@&!!!

How does our story end?









With great power comes
great responsibility.

Learn more at:
<https://matthewrenze.com>

