# Lecture 10: Fun & Games


## Matthew Caldwell


**COMP0088 Introduction to Machine Learning • UCL Computer Science**

# Contents

# 10.1: Caveat Emptor, Cave Canem

**COMP0088 Introduction to Machine Learning • UCL Computer Science**

# Forewarned is forearmed

- This week's topics will not be in the exam
- You may legitimately skip them if you are uninterested or short of time
  - → It's not as if you have a shortage of material to get to grips with already
- Some of them are fun and entertaining, but that doesn't mean they don't have serious consequences
- They are intended to add to your understanding of ML in real life
  - → But possibly that's not a realistic or useful ambition — only you can judge
- The treatment here will inevitably be extremely shallow
  - → You may have other modules that will cover some of this material in more depth

# 10.2: Reinforcement Learning

**COMP0088 Introduction to Machine Learning • UCL Computer Science**

# Neither supervised nor unsupervised be

- Reinforcement learning is often considered the third paradigm of ML, and its *modus operandi* is entirely different from the other two
  - → As mentioned way back in week 1, it also accords more with our intuitive understanding of the word learning
- For all of the ML models, algorithms and applications considered up to now, the training data is some external thing, an entity or corpus that we bring to the algorithm and serve up on a platter
- RL has to go out and get its own data: it must experiment and explore

# Markov decision processes

- The most common conceptual model underlying RL is the MDP
- MDPs are structurally similar to the HMMs we talked about a couple of weeks ago

  → Both are built on the foundation of Markov Chains

- They posit an observed system that occupies discrete states at discrete times, with stochastic transitions between states from timestep to timestep
- As with HMMs — and as the name suggests — the underlying process is assumed to be Markovian: the future is conditionally independent of the past, given the present

# Markov decision processes

- MDPs differ from HMMs in (at least) 3 important ways
- The state is not hidden
  - → We may not know all states (in RL, at least), be we can observe the current state
- We aren't just passive observers of the system, we get to interact with it
  - → At each time point we perform an action, and the combination of state and action determine the transition probabilities to the next state
  - → For "we" read some software agent — we won't get our hands dirty ourselves
- There may be rewards associated with each transition
  - → Rewards depend on actions as well as states
  - → Rewards accumulate over time and motivate the learning

# Policies, values and discounting

- An agent interacting with the system of an MDP chooses its actions based on a policy, π, which associates actions with states
  - → Policies may be deterministic: $a = \pi(s)$, essentially a lookup table
  - → Or probabilistic: $P(a|s) = \pi(a, s)$
- The overall reward for any individual episode (trajectory through the system) is the discounted sum of rewards for each of the individual time steps

$$R = \sum_t \gamma^t r_t$$

# Policies, values and discounting

- The discount factor $\gamma$ affects how much rewards are valued according to how far in the future they are received
  - → $\gamma = 1$ means future gains count just as much as present ones
  - → $\gamma = 0$ means it's now or never: we only care about instant gratification
- This matters because a policy might take a long time to pay off!
- The utility of a policy is the expected sum of the rewards from following it

$$U_\pi = \mathbb{E}(R_\pi) = \mathbb{E}\left[\sum_t \gamma^t r_t | \pi\right]$$

- Classically MDPs are used to model things like games (of skill and chance!) and investment strategies, or for optimising control processes
  - → In the latter case the rewards are typically costs and the goal will be minimisation rather than maximisation
- Given knowledge of the model parameters (transition probabilities and rewards/costs), an optimal policy can be determined via dynamic programming
- MDPs become problematic when the state and action spaces are large or the model structure and parameters are known incompletely or not at all

# Call for reinforcements!

- RL addresses incomplete or intractable MDP-type problems in two key ways:
  - → Sampling from the process to get data about how it works
  - → Function approximation (eg using deep neural networks) instead of explicit modelling to estimate the process behaviour: what are the states, transitions and rewards?

# State value function

- Define some auxiliary functions as stepping stones to the best policy
- $V_\pi(s)$ is the expected total reward for following policy π starting from state *s*

$$V_\pi(s) = \mathbb{E}\left[\sum_{t=0} \gamma^t r_t \;\middle|\; s_0 = s, \pi\right]$$

→ It tells us the value of being in that state

→ An optimal policy π* is one that maximises the value for every state

# Action value function

- $Q_\pi(s, a)$ is the expected total reward for taking action *a* from state *s* and then following policy π thereafter

$$Q_\pi(s,a) = \mathbb{E}\left[\sum_{t=0} \gamma^t r_t \;\middle|\; s_0 = s, a_0 = a, \pi\right]$$

  → It decomposes *V* for one immediate action, expressing the quality of that choice

  → The optimal policy π* will choose the action with highest *Q* at each state

- Obviously this way of defining the optimal policy is a bit circular

# Monte Carlo or bust

- Given a particular policy π, we can estimate $Q_\pi(s,a)$ by just trying out that choice a bunch of times and computing the average total reward obtained
  - → This and subsequent discussions presuppose that we have some way of actually running such tests — ie, a simulation environment
- In practice, runs will often be episodes of sequences of actions and rewards may not be known until the end, but you can associate terminal rewards with individual state-action pairs within the sequence
- If the state and action spaces are small, then it may be possible to find the best policy by trying all the actions for all the states enough times that you get a good *Q* estimate for every possible choice

# Bellman equation

- Because of the sequencing, *Q* can be divided up recursively like this

$$Q^{\star}(s,a) = \mathbb{E}\left[ r + \gamma \max_{a'} Q^{\star}(s',a') \;\middle|\; s,a \right]$$

- In an iterative learning scenario, this leads to an updating strategy like this

$$Q_{i+1}(s,a) = \mathbb{E}\left[ r + \gamma \max_{a'} Q_i(s',a') \;\middle|\; s,a \right]$$

→ ie, action value at each iteration depends on immediate reward plus previously estimated action value for all subsequent timesteps (appropriately discounted)

# Q-learning

- Explicitly learning $Q(s,a)$ scales badly with the number of states and actions
- Instead, we can substitute a function approximator such as a neural network

$$\tilde{Q}(s,a|\boldsymbol{\theta}) \approx Q^{\star}(s,a)$$

- This then learns to estimate $Q$ by updating its parameters $\boldsymbol{\theta}$

  → Using backprop and gradient descent in more or less the usual way

- If the function approximator is a deep neural network this is (surprise!) known as deep Q learning — and the NN as a deep Q network or DQN

# Exploration vs exploitation

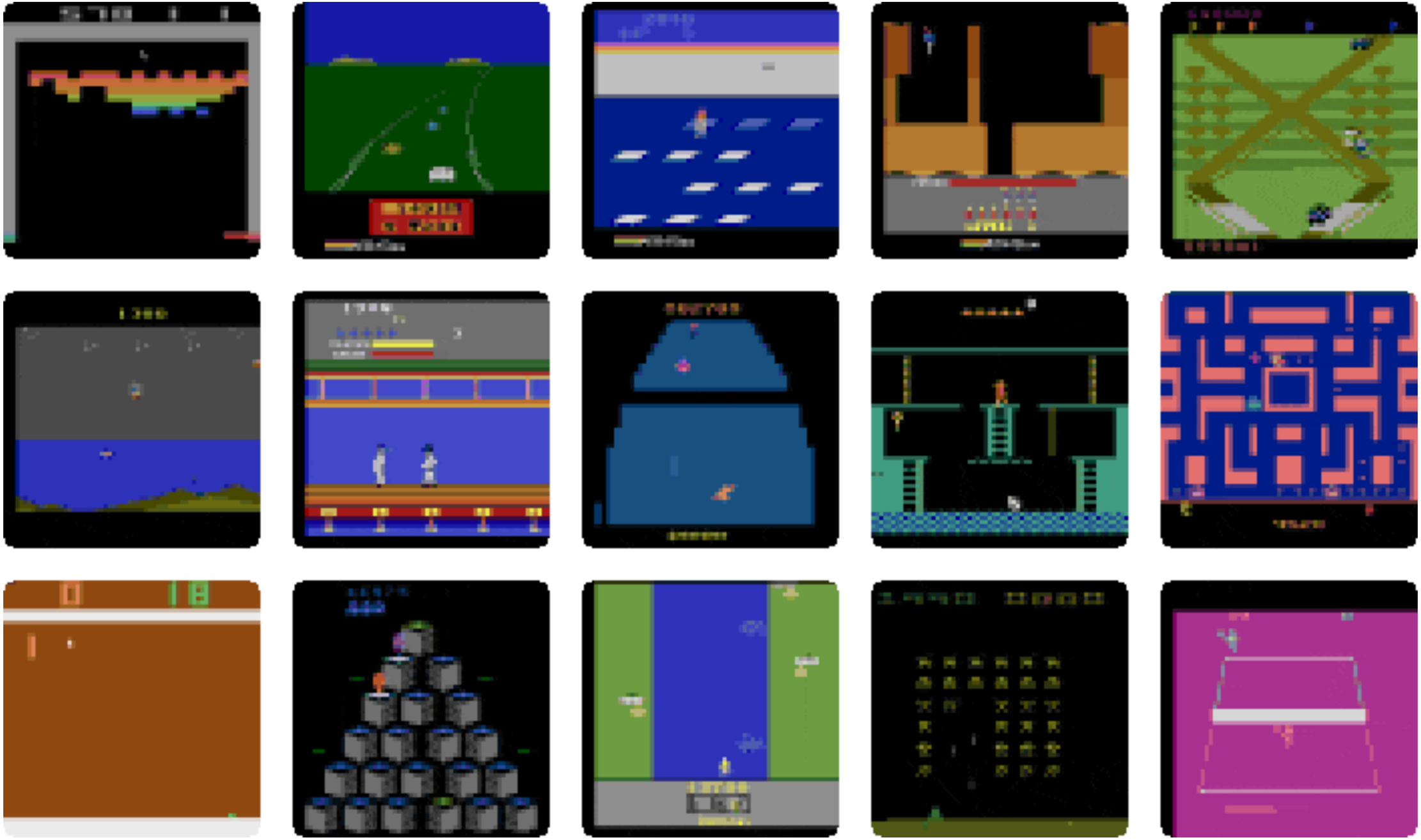
- An important notion for RL is the tradeoff between exploration and exploitation
  - → Trying out new actions or using what you already know
- Initially the agent will know nothing and has no choice but to explore
- But once it discovers some rewarding actions, it might just keep plugging away at those without ever bothering to look at other more complex paths that might ultimately lead to much higher rewards
  - → Conceptually similar to the idea of mode collapse in GANs

# Epsilon greediness

- A common approach to this is known as the epsilon-greedy algorithm
- A hyperparameter ε defines balance between taking optimal action at each step (according to your current knowledge) or just picking at random
- This ensures that there's always some chance of doing something novel even when you already have some profitable strategies
- Some variations allow this to vary throughout training — eg, explore a lot at the start, exploit more later

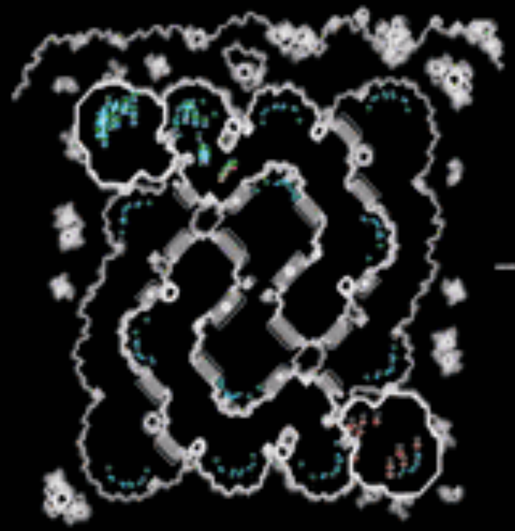
# Deep Q greatest hits

- Atari 2600 games
- Starcraft
- AlphaGO (not purely RL — bootstraps off a lot of expert domain knowledge — but also uses a huge amount of self-play RL)

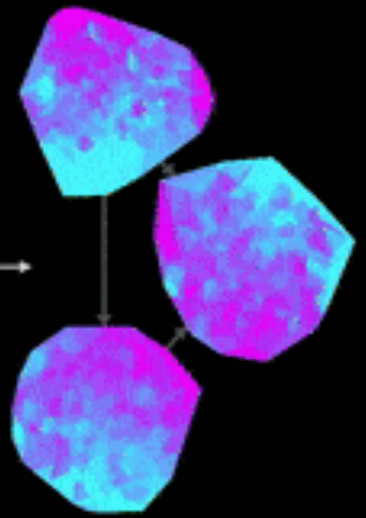




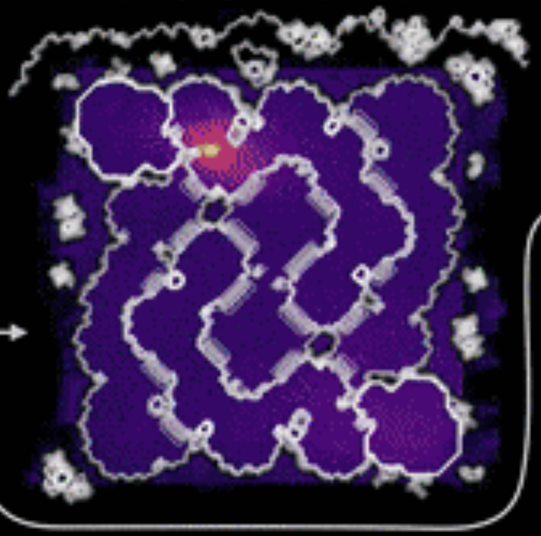
Raw Observations



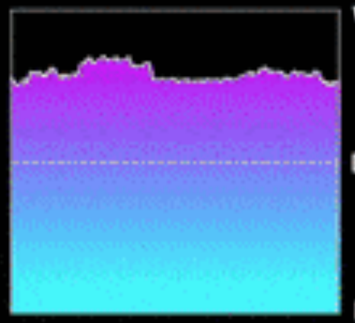
Neural Network Activations



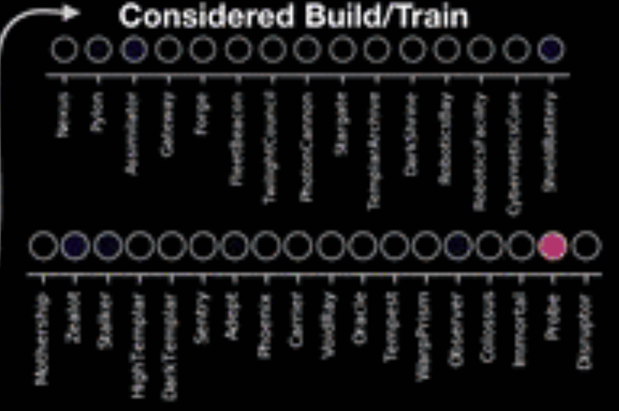
Considered Location

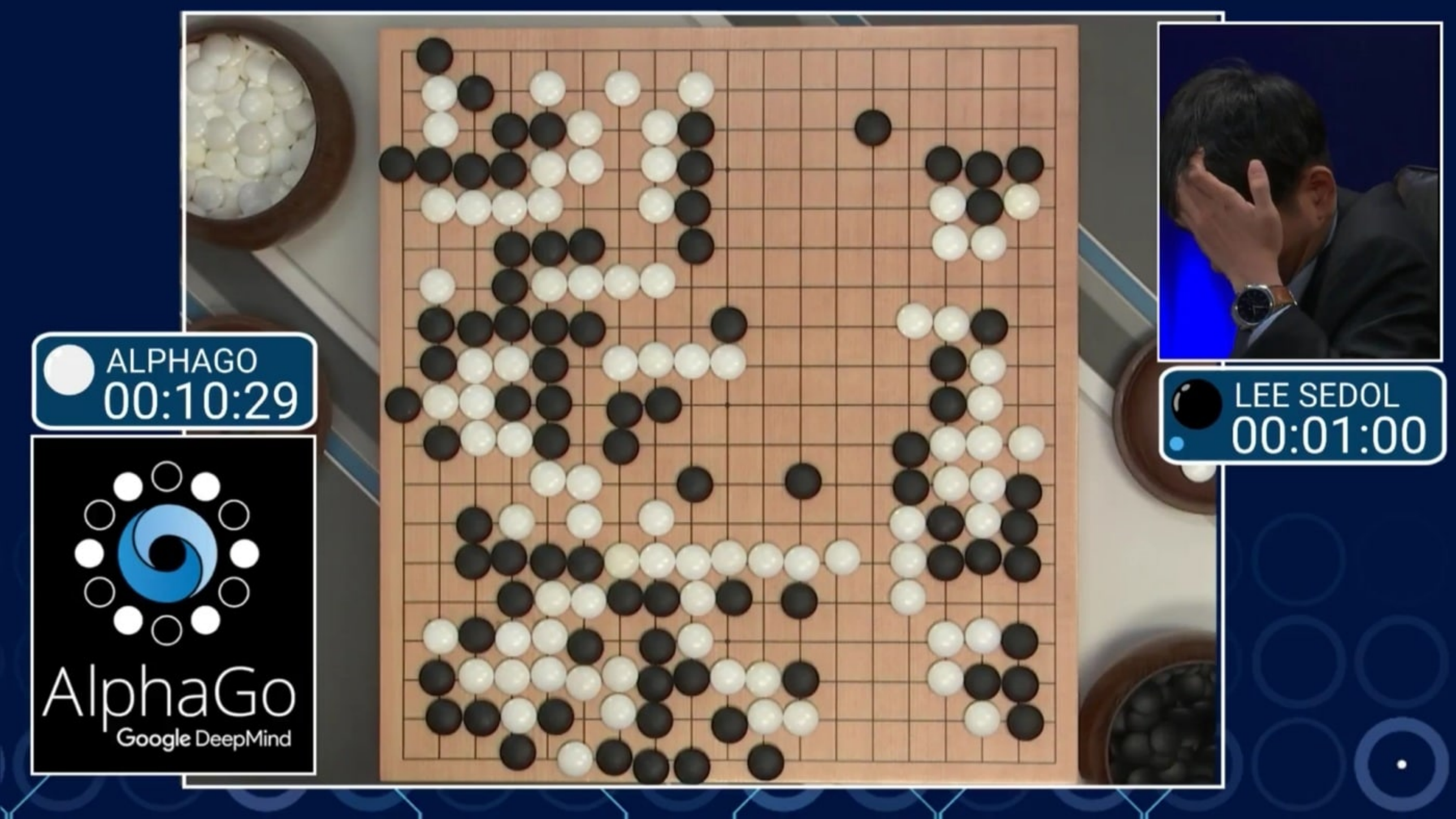


Outcome Prediction




Considered Build/Train




ALPHAGO
00:10:29
AlphaGo
Google DeepMind
LEE SEDOL
00:01:00

# Reinforcement learning wrap up

- RL addresses a whole range of problems for which supervised and unsupervised approaches basically have nothing to offer: dynamic and interactive problems where static data can’t readily capture the whole story
- Requires an environment that the agents can explore a lot — which usually means a simulation rather than real life
- It can be complex and expensive and unstable to train
- Companies like DeepMind have invested very heavily in RL approaches and it seems to have a rosy future

# 10.3: Deepfakes

**COMP0088 Introduction to Machine Learning · UCL Computer Science**

# Synthetic media revisited

- As discussed last week, generative models such as GANs and VAE attempt to capture the distribution of real data — most often images — in such a way that new samples can be drawn from the distribution
- These samples do not directly correspond to any of the training images
  - → Although they are in some way a combination of them
- Large models such as BigGAN and StyleGAN are capable of generating convincing synthetic images of people and scenes that have never existed
- If one were interested, and idle, and maybe just a teensy bit ethically compromised, what might one do with this sort of technology?

# Rise of the deepfakes

- CGI has been around for decades, becoming ubiquitous in movies and TV
- Simulacra of real people have increasingly turned up in Hollywood fare in recent years, eg Peter Cushing & young Carrie Fisher in Rogue One
- Many standard kinds of media manipulation tech have long been available to the general public via applications like Photoshop, After Effects etc
- A number of computer vision and machine learning techniques have been incorporated into these behind the scenes for things like denoising
- But public awareness of synthetic media with a substantial deep learning component dates to late 2017, when user **deepfakes** posted some face-swapped porn videos to Reddit

- Similar face swapping abilities and video augmentation methods have become increasingly widely available since
- Many silly, low-quality examples crop up as novelty apps that turn a photo into a campy lipsync of I Will Survive or whatever
- These manipulations are in many ways very impressive, but they're fooling no-one — are intended to fool non-one — and are mostly quite peripheral to deepfake hysteria

  → Even though they are to a large extent the same thing

# The art of collage

- Deepfakes are basically never completely synthetic
  - → Then again, neither are any of the other GAN outputs — they are always in some way a smoothed patchwork over the training set, albeit constructed at one remove
- People — especially the people who inhabit sleazy fake porn forums — want to see fake Scarlett Johansson or whoever, not some non-existent non-entity
  - → Even if that fake Scarlett Johansson is *terrible*, little more than a shop window mannequin in a Black Widow mask
- Also, these kind of deepfakes *are* mostly terrible, a far cry from the slick media offerings that are the public face of deepfakes

# The public face of deepfakes

- Buzzfeed & Jordan Peele: Barack Obama
- Channel 4: Alternative HMQ Xmas message
- Tom Cruise TikTok


TikTok
@deeptomcruise


TikTok
@deeptomcruise


los angeles
at noon
TikTo
@deeptomcrui

# The public face of deepfakes

- These were all explicitly promoted as warnings against the looming threat of the infocalypse, when everything is possible and nothing makes sense
- Nothing can be trusted anymore, your senses are lying to you, infamy, infamy!
- But also, y’know, watch our channel! Visit our advertisers! Don’t forget to like and subscribe!

# Gatekeepers and moral panic

- The most alarmist takes on deepfake synthetic media come from the people and organisations whose self-images and business models are predicated on their superior access to — and gatekeeping of — the media in question
  - → We saw a similar reaction from primarily text-based media organisations to GPT-3
- And the most prominent examples, presented as warnings of the imminent threat of deepfakes, come from those same organisations or people
  - → I absolutely am not trying to encourage conspiratorial thinking about the “MSM” — creators and platforms that play important roles in our society — but the reflexive alarmism does at least suggest some unease with the potential loss of privileged access to — or control of — the production process

# Fear sells

- First and foremost, these high end deepfakes are entertainments, products of the financial clout and expertise of media companies — and parts of their workforce with a very particular set of skills
  - → or portfolio pieces for said workers, in the Tom Cruise case
  - → a significant fraction of deepfake technology is arguably just an arm of VFX
- Scare stories about the infocalypse get bums on seats
- And if a side effect is to sow suspicion about the democratisation of media creation as some kind of conduit for terrorism, well who's complaining?

# Trust no-one

- But self-serving scaremongering about deepfakes is not harmless
- Trust in politics, journalism, media and institutions is extremely low
- Anti-establishment sentiment and conspiratorial thinking has become a hallmark of political campaigning, including in the UK referendums on Scottish independence and Brexit, and in recent US presidential elections
- Donald Trump explicitly campaigned and governed on a platform of mistrust, denouncing everything as fake news as a matter of policy
- Arguably the threat of deepfakes lies at least as much in the fomenting of an atmosphere of generalised mistrust and a poisoning of public discourse, as in the potential harm from any particular instance of deepfakery

# Deepfakes and cheap fakes

- As pointed out at length by Britt Paris & Joan Donovan in their influential paper for Data & Society, fake media is nothing new
  - → The camera has always lied
- Most of the techniques employed by scammers and propagandists remain at the cheap fake end of the spectrum
- There is certainly some technological democratisation in the era of Photoshop and Instagram
- But the balance of power between those with access to media-faking capabilities and those victimised by them remains overwhelmingly the same as it ever was: if Rupert Murdoch wants to ruin your life, it will be ruined. If you want to ruin his, not so much.

- This is not to say deepfakes themselves are harmless
- Victims of deepfake revenge porn or other kinds of reputational damage may suffer a lot of hurt and hardship, loss of livelihood, broken relationships, disgrace, perhaps even as a consequence injury and death
- Deepfakes may present novel opportunities for nasty crimes, but they are not really novel crimes

  → It's mostly the same kinds of perpetrators targeting the same kinds of victims
- And to a large extent the evidence suggests that the supposed convincingness of the fakery, the headline-grabbing technological magic of AI, is in many cases a red herring

# Does authenticity even matter?

- People are often predisposed to see what they want to see and ignore what they'd prefer not to
- Tribal loyalty trumps the act of seeing with one's own eyes
- Cognitive dissonance is a helluva drug
- Some of the most successful, most widely-shared and readily believed examples of post-truth put little or no effort into realism
  - → Sharpiegate is a notorious case in point
- Beyond a certain point, all that matters is the claim — any ostensible evidence for the claim is just a decorative flourish, if that


MS
AL
GA
8 AM Tue
8 AM Sun
8 AM S
FL
H
M
M
M
Bahamas
8 AM Mon
Cuba

# The infocalypse will not be televised

- There are all sorts of reasons to think the human race might be in trouble
- There are a lot of us, and we're not psychologically equipped to deal with that fact, let alone be acquainted with so many other people's thoughts
- We are credulous and arrogant and territorial and irrational and cruel
- There's a good chance we'll go to hell in a handcart packed full of denialists and conspiracy theorists demanding we go faster
- But it doesn't seem very likely that deepfakes will have a significant impact on that outcome one way or another
- As always, people are the problem

# 10.4: Ethics & Fairness

**COMP0088 Introduction to Machine Learning · UCL Computer Science**

# What could possibly go wrong?

- We've mostly — give or take the odd snarky aside — assumed that all of the learning algorithms and models discussed are morally neutral
- But no technology exists in a vacuum, whether that technology is saving lives or destroying livelihoods
- And given the breadth of applicability of ML and the depth of data and assumptions feeding into those applications, it behoves us to consider what the effects of this stuff might be, and how complicit we are in those effects

# Garbage in, garbage out

- Let’s start with the data, which is perhaps the least contentious element
- As discussed a number of times before, ML is always and inevitably shaped by the data used for training and validation
- If you train on bad data you can’t be surprised to get bad results
  - → And a lot of data is bad, one way or another
- But how do you know? We rarely have the luxury of carefully curating our own datasets
- And even with detailed individual curation you’re likely overestimating your own objectivity and encoding your own unexamined assumptions about what matters and what doesn’t

# Unconscious bias strikes again

- These biases are almost by definition unintentional but just because you don’t notice them that doesn’t mean they’re not there
- Choices in data and assumptions in modelling inevitably reflect the worldview of those making them
- Often people are casually blind to their prejudices, unaware of how their attitudes are being consistently reinforced by the monoculture they inhabit
- Tech education, research and business was overwhelmingly the province of middle class white men for so long you could easily mistake that for the natural order of things, if you didn’t look too closely

# Lena

- For decades the image of Playboy centrefold Lena Forsén was a *de facto* standard test case in image processing and computer vision
- This usage was not ill-intentioned, just unthinking
- But the implicit message was: this scientific discipline, like all scientific disciplines by divine right, is the province of leering straight men
  - → And what sort of humourless, man-hating termagant could possibly object to such a harmless bit of fun?
- This isn't training data — it's much more visible than that — a flagship
- Imagine how much more easily such unexamined biases can lurk below the surface, hidden in the data morass

# Big data, big trouble

- The bigger the dataset, the more its collection may be out of our control
- It’s likely to involve either automation or a long history, or both
- Either way, there will probably be plenty of unexamined assumptions
- We can try to do some filtering and curation along the way, but this is increasingly difficult at scale

  → And of course is just another route for embedding our unconscious biases

# Check your working

- There's a good chance we might only find out what's in our data after we fit our model and see what it does
- This is one reason why testing and analysis are so important — but this requires intentionality
- If it doesn't occur to you to look for biased and artefactual behaviour in your models, you probably aren't going to find them
  - → Maybe someone else will, which might be potentially embarrassing
- Shoutout again to ConceptNet-Numberbatch for actively doing this kind of thing

# Dubious labels

- Data for supervised learning requires labels, which means someone or something has to provide those
- For small datasets we might do it ourselves
- For larger ones there are two other common choices: crowdworking and automation
- In both cases there are issues of quality control
- Crowdworking practices may be exploitative and draw from unrepresentative populations
- They are inevitably subject to the vagaries of their assembled multitudes

# Self-fulfilling fallacies

- There are several approaches to quality control for crowd-sourced labels
- There will be a replication requirement — samples must be independently labelled by multiple workers
- In addition there may be automated monitoring
- But this gives rise to problematic feedback loops
- In particular, labels may be checked against automated classifications and unexpected results weeded out
- But this just amplifies any biases already inherent in the model
- Effectively the crowdworkers are required to rubber-stamp judgements already encoded, adding unearned legitimacy

# Ownership and consent

- Historically, researchers and companies have been extremely cavalier about gathering training data
  - → Whatever you can scrape together is great; let's not get hung up on the details
  - → The explosive growth of the internet produced a free-for-all data bonanza and every <img> tag was fair game
- There was an unspoken assumption that whatever remnants of data were left in a trained model were so abstracted from the original as to be untraceable
- With the rise of very large deep learning models like GPT-3 and BigGAN, this is clearly no longer the case

- Data generated by these models may be novel in one sense, but it's also very clearly a gloss on the original training data
- These models learn distributions, but they do so via the medium of samples
- Are they plagiarising the training data?
- Are they appropriating it?
- If some future GPT-*n* can churn out new but stylistically unmistakeable Stephen King bestsellers every hour, what does that do for his market value?
- If StyleGAN captures what it knows of a particular facial feature from photos of you, does it owe you something for that? Is it stealing your soul?

# Tasks

- Once you have the data, the question remains: what are you doing with it?
- Do the problems you’re attempting to solve make sense?
- Is your approach likely to produce valid answers?
- And how will those answers be used?

# The lure of phrenology

- If you ask a stupid question, you should expect a stupid answer
- But stupid questions can be awfully tempting
- People love spurious patterns and comforting explanations
- We like to get an edge, to have special insight, to be in on a secret
- ML can provide useful tools for navigating data, but it is also very well suited to noise mining, to the discovery of imaginary relationships

  → This is why it's so important to be on guard against overfitting

# Fish and ye shall find

- Fishing expeditions abound
- It is all too easy to cast specious overfits as profound discoveries
- To overstate the significance of results that accord with what you want to show
- Never trust a p-value: significance tests are the last refuge of the scoundrel
- Bandying around irrelevant metrics for marketing purposes is commercially commonplace, but p-hacking is notorious in science as well

# Dangerous visions

- Whether or not the patterns you claim to detect are real, there's also the question of what will be done with them. What is your model for?
- Will discoveries of financial patterns lead to instability or inequality?
- Will medical diagnostics lead to better treatments or to marginalisation and exclusion of specific patient classes?
- What does it mean to identify — or claim to identify — traits that may be seen as socially undesirable, such criminality or minority sexuality?
- How much are such labels actually proxies for existing structures of inequality and disenfranchisement? How much do they just codify prejudices and furnish them with a veneer of objectivity?

# Safeguarding

- Is it safe to deploy your model? Who is ill-served by it?
- In a few contexts — notably medicine — there are regulatory frameworks that demand rigorous testing — but even there errors can creep in with potentially dire consequences
- Elsewhere it's the wild west. Most ML applications are self-policed at best, and regulators, if they exist at all, are ill-equipped to monitor
- There are always incentives to cut corners, to oversell capabilities and understate problems
- Of course lots of applications are frivolous and trivial, with no apparent consequences, but how do you know that your mickey mouse model isn't paving the way for some future dystopian use case?

- Businesses look to ML for insight, automation, efficiency
- To improve their processes, cut their costs, increase profits and outfox competitors
- Governments employ ML for all those reasons — nations are corporate bodies too — and also for intelligence, security, social control
- All of these can involve ethical conundrums: what social norms is it reasonable to enforce? At what point does the urge to increase engagement give way to the need to avoid radicalising your viewer base? Does corporate responsibility even mean anything?

# Pitchforks & torches

- These things are not in our control and maybe not in anyone’s by now
- YouTube and Facebook are entrenched enough to be laws unto themselves
- And perhaps it doesn’t matter in the end, perhaps we’ll all he happy to sleepwalk into some deep learning utopia ahead, when the techbros take on climate change and win
- Or perhaps there’ll be a tectonic shift in public perceptions and mobs of villagers with pitchforks and torches will storm Silicon Valley — or Gower Street
- Perhaps it’s too late to do anything about that
- All we can really do is try to act thoughtfully and deliberately and ethically on our own tiny patches of the ML landscape

# 10.5: Outro

**COMP0088 Introduction to Machine Learning · UCL Computer Science**

# Scratching the surface

- ML encompasses a plethora of techniques, from the simplest least-squares regression or nearest neighbours classification to huge generative models capable of writing at least superficially-convincing text and hallucinating seemingly photorealistic images of non-existent people doing non-existent things in non-existent places
- Some of these require painstakingly labelled data, some learn approximate input distributions, yet others can probe simulated environments to figure out how they behave
- We have only scratched the surface in this module, but hopefully enough to whet your appetite and give you a toehold, a base of understanding on which to build in the future

# Ubiquity & utility redux

- ML applications saturate our lives and increasingly mediate our interactions with the world
- They give us tools to navigate an increasingly unwieldy information environment, and also give others the tools to interrogate and manipulate our interactions with that world of data
- This tendency is only going to increase

# Goodnight & good luck

- I hope you've enjoyed the module

  → I have!

- I hope you'll find ways to use your ML knowledge productively
- I hope it can help you understand the ways others are using theirs
- The rest is up to you