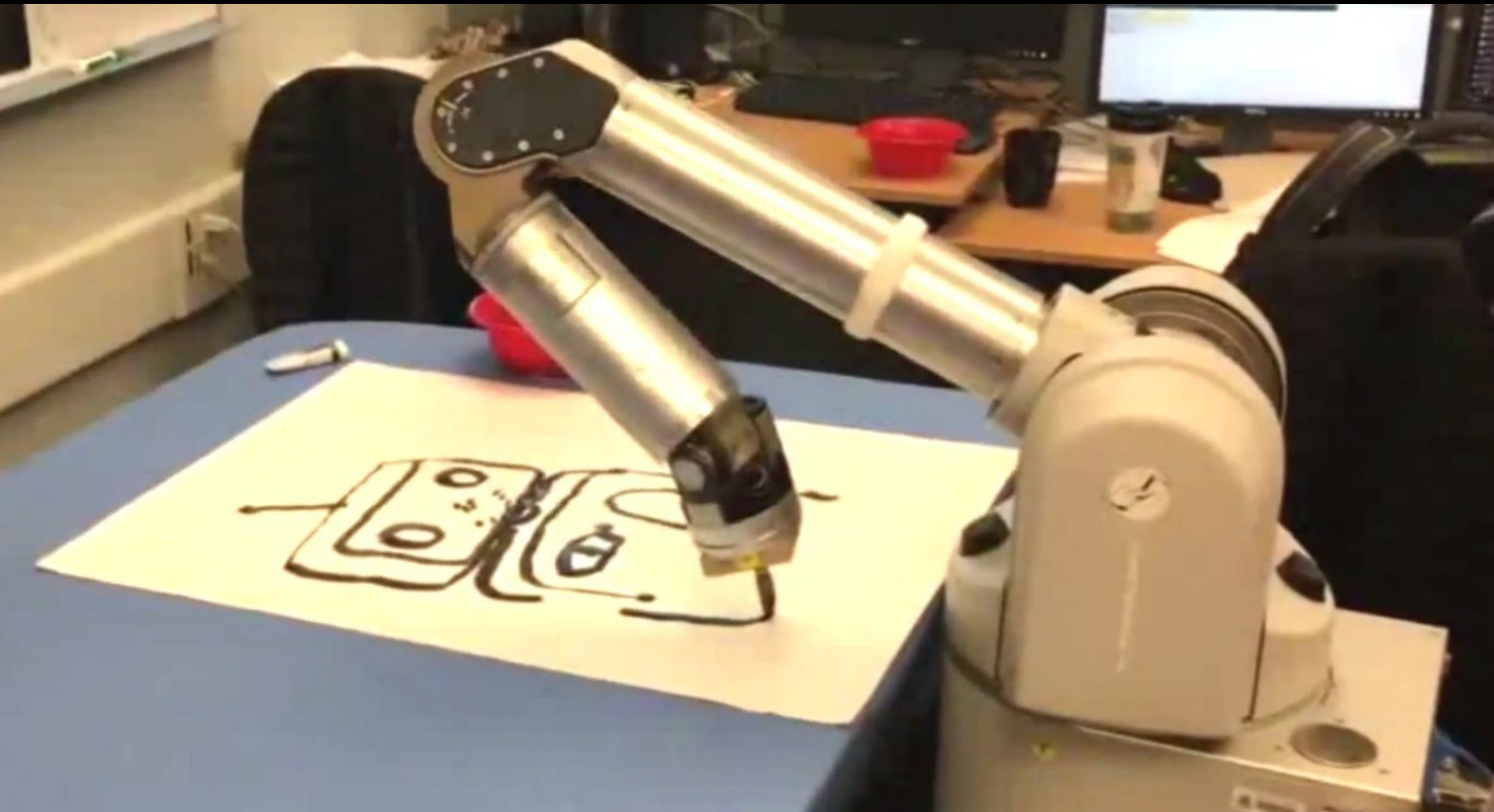


ROBOTS BEHAVING BADLY:
COPYRIGHT LIABILITY IN THE AGE
OF ARTIFICIAL INTELLIGENCE

DR ANDRES GUADAMUZ, UNIVERSITY OF SUSSEX

PREVIOUSLY, IN CODAIP...

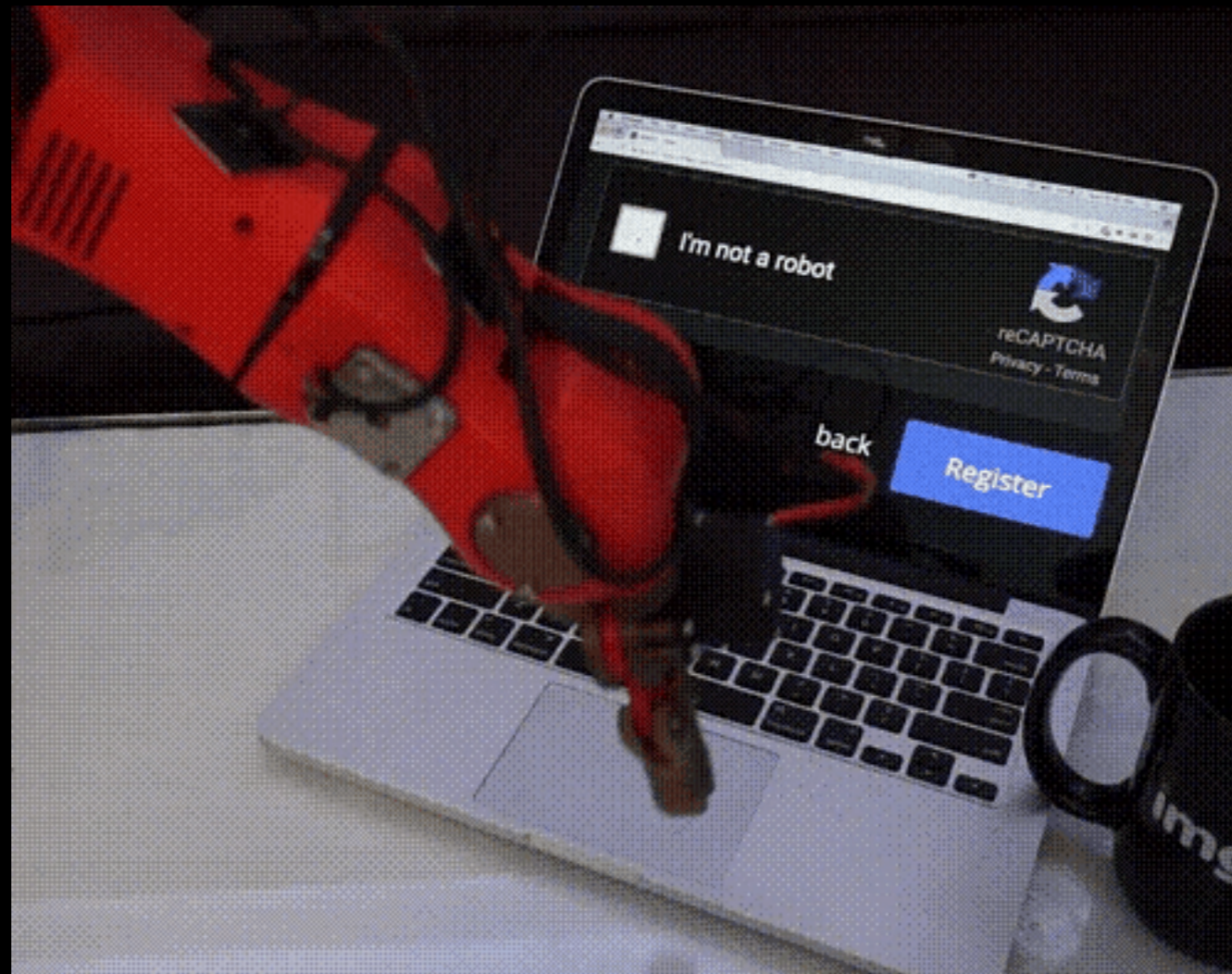


OPTIONS FOR AI COPYRIGHT

- No copyright due to no originality/creativity.
- No registration.
- Make UK's approach more widely used, programmer gets copyright.
- Take a wider approach: "copyright laws . . . do not expressly require 'human' authorship." *Urantia Foundation v. Maaherra* (1997).
- Artificial Intelligence rights?



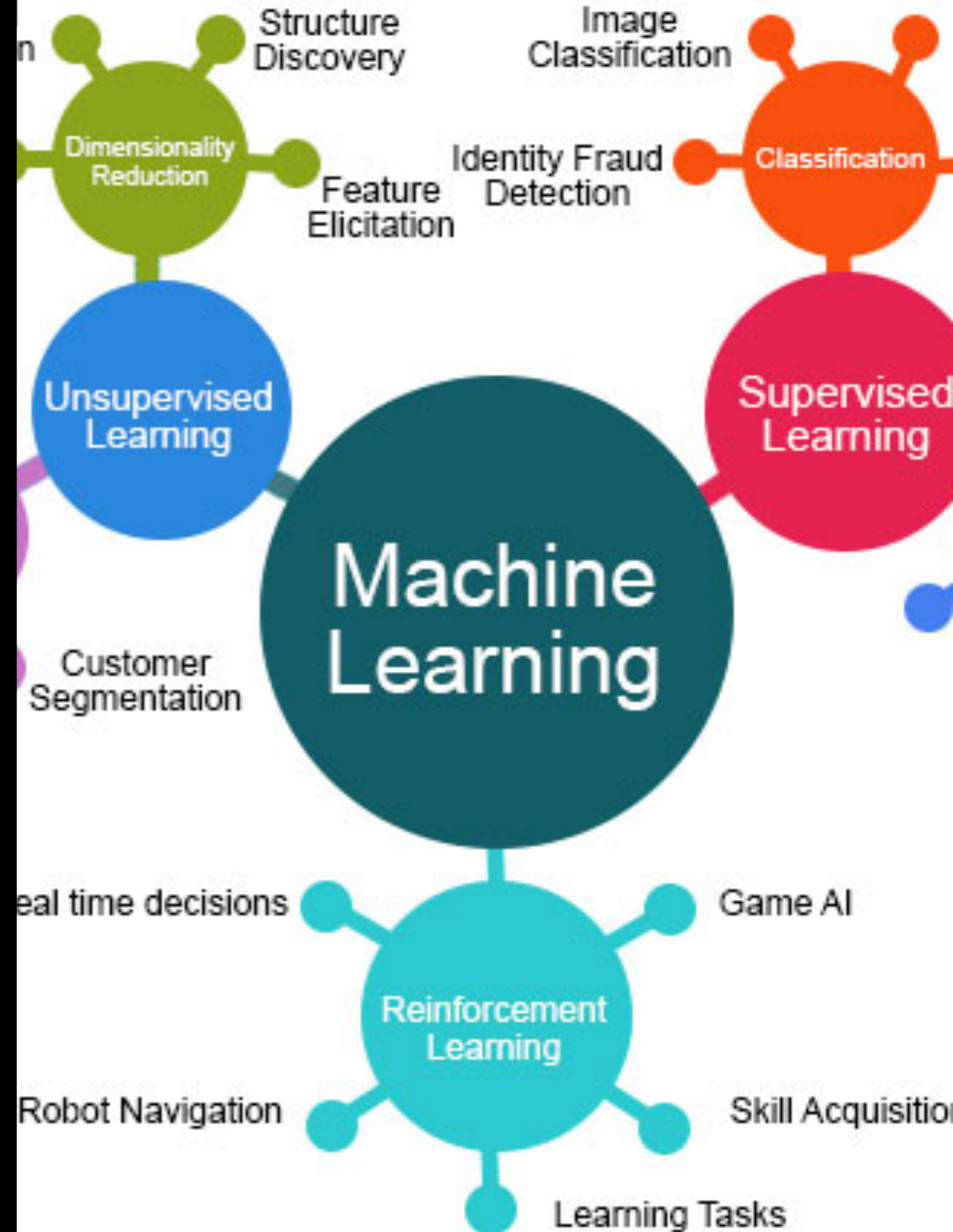
AN APOLOGY



KEY CONCEPTS IN AI AND CREATIVITY

- Algorithm
- Software
- Data
- Datasets
- Output

Machine Learning Algorithms



FRAMING THE PROBLEM OF INFRINGEMENT

1. An autonomous agent infringes copyright on its own accord.
2. Dataset infringes copyright (exceptions and limitations).
3. Derivative work could infringe copyright.



AUTONOMOUS AGENTS

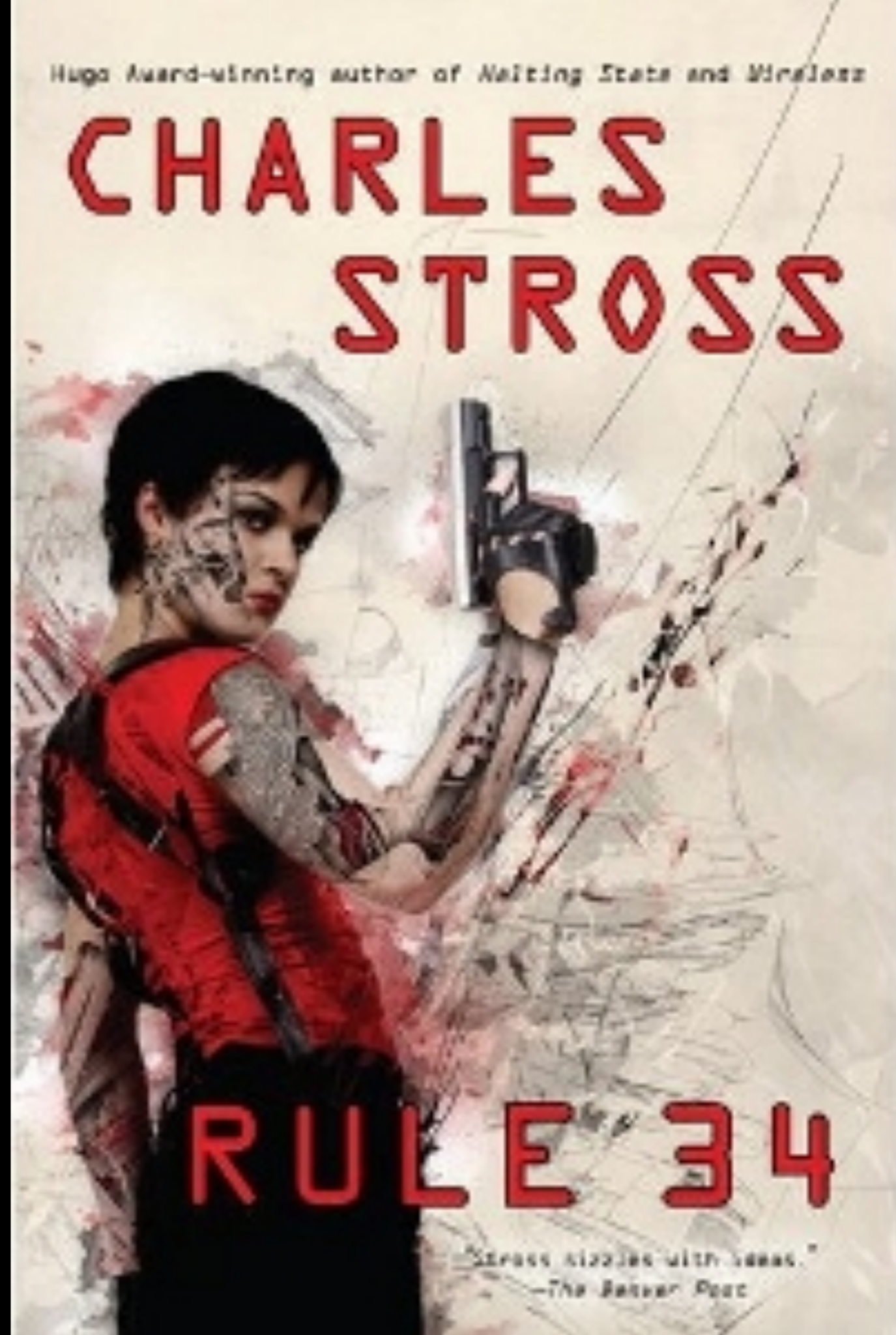


WHAT PEOPLE THINK WHEN YOU MENTION AI LIABILITY



WHAT GOT ME INTERESTED (IN FICTION)

- Rule 34 by Charles Stross (sequel to Halting State).
- Book describes a rogue AI that starts killing people with IoT devices.
- Led to exploration of liability of AI.



!MEDIENGRUPPE BITNIK

SHOPPING BOT



! MEDIENGRUPPE BITNIK

- Random Darknet Shopper
- Botnet buying random items from the Darknet using Bitcoins.
- It purchased drugs.
- Police confiscated the bot, then released.
- Public prosecutor deemed the artistic work outweighed any possible damage of purchasing drugs.



PROSPECT FOR COPYRIGHT INFRINGEMENT

- This is not new, we've had spiders scraping data off the Internet for decades.
- It may depend on what the AI is doing.
- Also related to text and data mining.



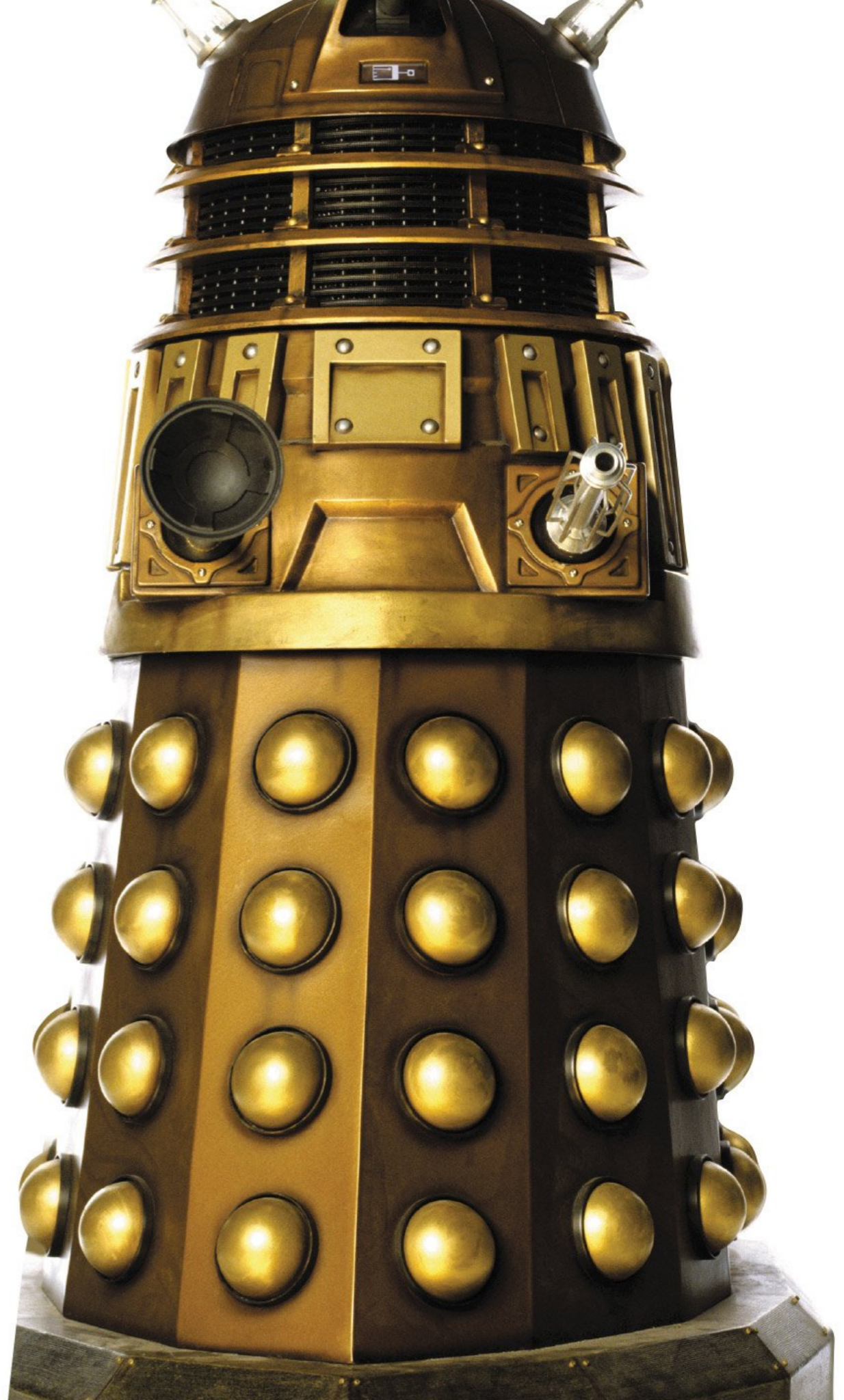
TEMPORARY COPIES

- Art 5(1) Directive Directive 2001/29/EC. Temporary copies are exempted if they:
- 1) **constitute an integral and essential part of a technological process;**
- 2) pursue a sole purpose, namely to enable the **lawful use** of a protected work; and
- 3) do not have an **independent economic significance** provided that:
 - 3.1) the implementation of those acts **does not enable the generation of an additional profit going beyond that derived from the lawful use** of the protected work;



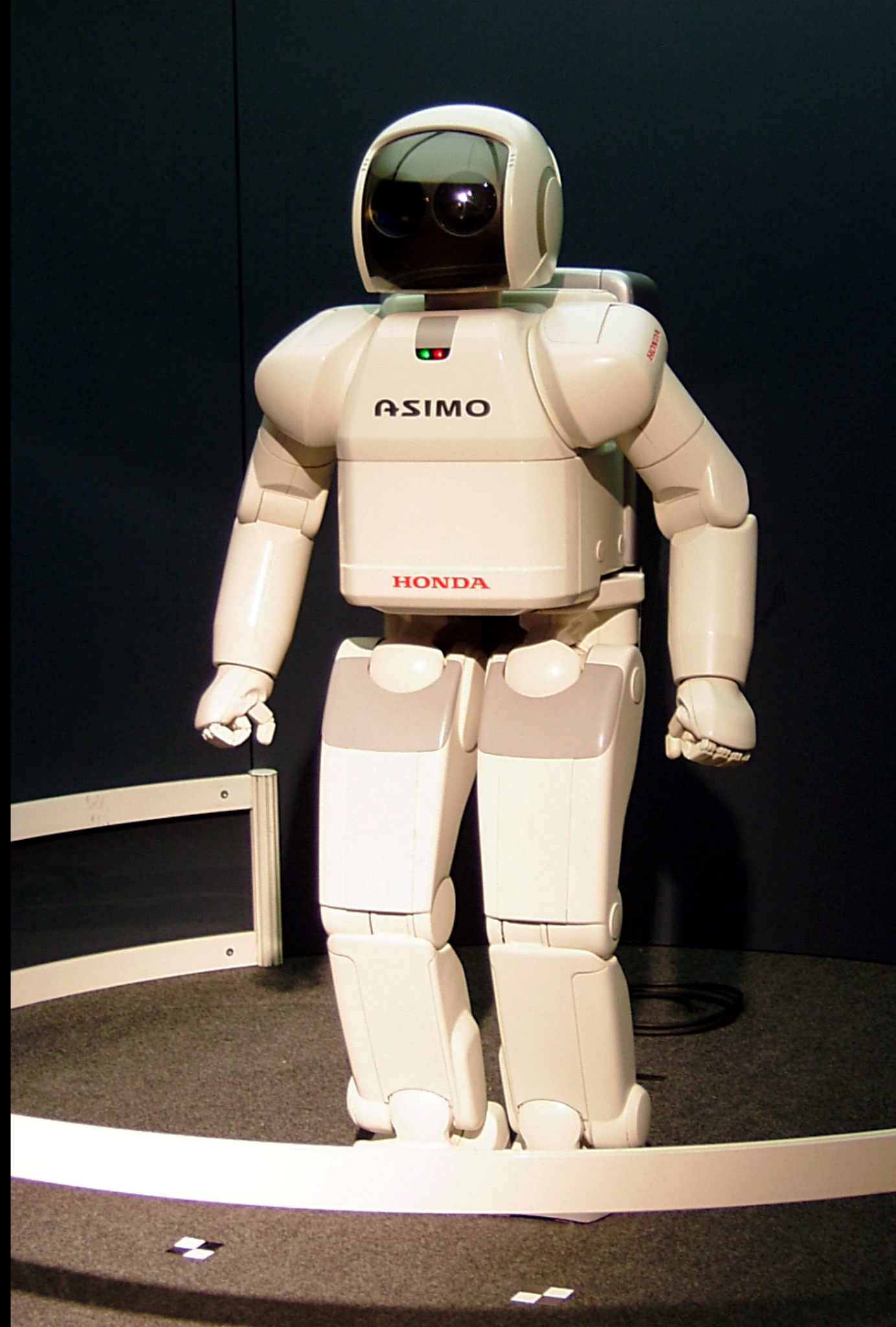
UK 2014 TEXT AND DATA MINING EXCEPTION

- Part of wide-ranging implementation of new exceptions to fair dealing list, including parody and private copying.
- s 29A CDPA. Exception for making a copy by a person who has lawful access for the sole purpose of research for a non-commercial purpose.
- Must be accompanied by sufficient acknowledgement.



JAPANESE EXCEPTION

- “For the purpose of information analysis (‘information analysis’ means to extract information, concerned with languages, sounds, images or other elements constituting such information, from many works or other such information, and to make a comparison, a classification or other statistical analysis of such information; the same shall apply hereinafter in this Article) by using a computer, it shall be **permissible to make recording** on a memory, or to make adaptation (including a recording of a derivative work created by such adaptation), of a work, **to the extent deemed necessary(...)**”



DSM DIRECTIVE (EU) 2019/790

- Art. 3: exception for "... reproductions and extractions made by research organisations and cultural heritage institutions in order to carry out, for the purposes of scientific research, text and data mining of works or other subject matter to which they have lawful access."
- Art 4: Exception for "... reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining."



FAIR USE IN THE US

- Google's unauthorized digitizing of copyright-protected works, creation of a search functionality, and display of snippets from those works are non-infringing fair uses. The purpose of the copying is highly transformative, the public display of text is limited, and the revelations do not provide a significant market substitute for the protected aspects of the originals. Google's commercial nature and profit motivation do not justify denial of fair use.
- Google's provision of digitized copies to the libraries that supplied the books, on the understanding that the libraries will use the copies in a manner consistent with the copyright law, also does not constitute infringement.

Google books

Google has reached a [groundbreak](#)

Fiction

[Literature](#)
[Science fiction](#)
[Fantasy](#)
[Romance](#)
[Mystery](#)
[Fairy tales](#)
[Short stories](#)
[Poetry](#)

Non-fiction

[Philosophy](#)
[Economics](#)
[Political science](#)
[Linguistics](#)
[Mathematics](#)
[Physics](#)
[Chemistry](#)
[Biology](#)

Random subjects

[Explanation](#)
[Epic poetry](#)
[Irish folklore](#)
[Poetry](#)
[Drama](#)
[Ballads](#)

Interesting



Classics



SOME CONSIDERATIONS

- This infringement will likely fall under many existing limitations.
- Temporary copies not covered as infringement, and most TDM could fall under this.
- Cumulative copying may not amount to infringement, no substantive copying.
- Revisit the Roman law of slavery? *Servus non habet personam*.



DATASET INFRINGEMENT



WHAT ARE THE ISSUES?

1. You need to have access to vast amounts of works in order to train the AI, this could infringe copyright.
2. Can the copyright owner of one such work sue the AI creator for hosting and analysing large amounts of data?



ACCESS TO DATA

- Use public domain works.
- Use works with permission, under some sort of licence.
- Rely on exceptions (temporary copy, data mining).
- Argue that you are actually not infringing copyright.



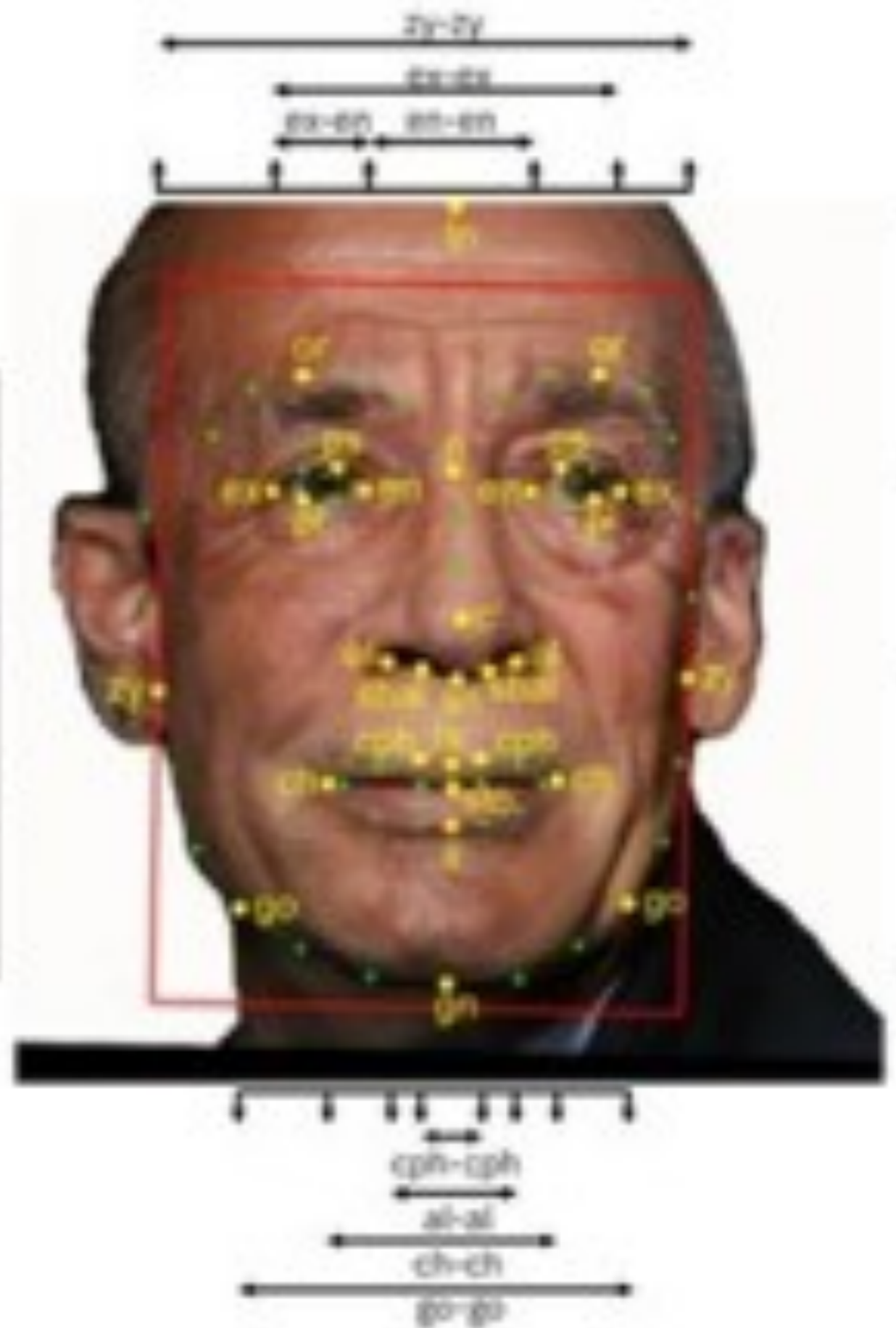
CASE STUDY 1

DIVERSITY IN FACES









DIVERSITY IN FACES

- Diversity in Faces is a dataset curated by IBM, selected from a Yahoo image dataset taken from CC-licensed content on Flickr.
- Dataset focuses specifically in "diverse" faces to train face recognition AI against bias.
- Images stored by IBM in identifiable and downloadable format.



DIVERSITY IN FACES

- Large majority of images shared with non-commercial licenses.
- If dataset is used for commercial purposes, it could be in breach of licence, hence inclusion would infringe copyright.
- No legal action (yet).

<i>License</i>	<i>Photos</i>
	17,210,144
	9,408,154
	4,910,766
	12,674,885
	28,776,835
	26,225,780
Total	99,206,564

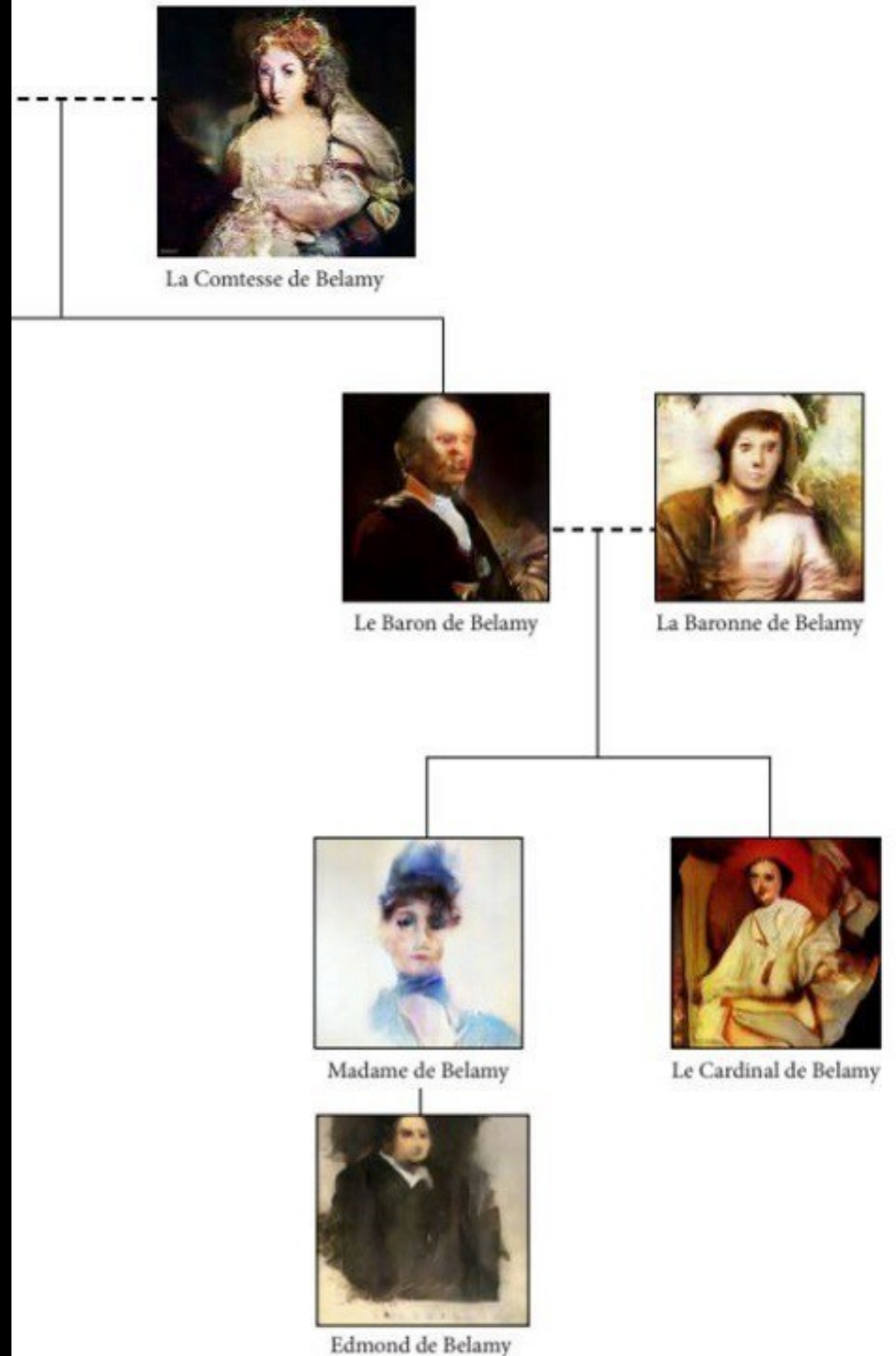
CASE STUDY 2

OBVIOUS ART



OBVIOUS ART

- Obvious Art used a Generative Adversarial Network (GAN) to produce their family of portraits.
- Original GAN algorithm was made by Ian Goodfellow, and released as open source software.
- Robbie Barrat, a researcher in Stanford, used portraits to train his version of GAN, he released the works under an open source software licence.



OBVIOUS ART

- Confusion as the various types of protection involved.
- Algorithms not protected by copyright, software code that runs an algorithm carries copyright.
- Dataset composed of public domain images.
- Data output (the portraits) may not have copyright in Europe.

The screenshot shows the GitHub interface for the repository 'robbiebarrat / art-DCGAN'. At the top, there are navigation links: 'Why GitHub?', 'Enterprise', 'Explore', and 'Marketplace'. Below the repository name, there are tabs for 'Code', 'Issues 11', 'Pull requests 2', and 'Security'. A light blue banner with the text 'Join' and 'GitHub is home to over 40 million developers...' is visible. The repository description reads: 'Modified implementation of DCGAN focused on generative art and others.' Below this, there are tags: 'gan', 'generative-art', 'art', and 'painting'. Further down, it shows '41 commits' and '1 branch'. At the bottom, there are buttons for 'Branch: master' and 'New pull request', and a commit history entry for 'robbiebarrat Update LICENSE.md' with a folder icon for 'checkpoints' and the text 'Replaced keepempty.py with'.

CASE STUDY 4

GPT-2 PREDICTIVE TEXT TOOL



GPT-2 Explorer

This demonstration uses the public **345M** 117M parameter **OpenAI GPT-2** language model to generate sentences.

Enter some initial text and the model will generate the most likely next words. You can click on one of those words to choose it and continue or just keep typing. Click the left arrow at the bottom to undo your last choice.

Sentence:

Andres is a member of the Spanish National Assembly and a member of the European Parliament.

Options:

- 39.1% **He**
- 19.7% ↵
- 7.0% **<|endoftext|>**
- 4.1% **His**
- 3.5% **In**
- 2.4% **The**
- 1.3% ↵↵
- 0.8% **A**
- 0.8% **As**
- 0.6% **Since**
- ← Undo

GPT-2

This demonstration uses the public **345M** 117M parameters

Enter some initial text and the model will generate the most likely continuation. You can choose it and continue or just keep typing. Click the left arrow to regenerate.

Sentence:

```
Andres is a member of the Spanish  
National Assembly and a member of the  
European Parliament.
```

- A tool by OpenAI that matches future text. What sets it apart is the quality of training content, they took Reddit links with at least 3 karma as an indicator of quality.
- The full index of scraped text is available for download.
- Is this copyright infringement?

CASE STUDY 4

FOLKRNN (AKA BOT DYLAN)

⏮ ⏪ 0:42 120 TEMPO ⏩ ⏭

SEE IT

The image displays four staves of musical notation in 4/4 time. The notation is written on a treble clef staff. The melody consists of various note values including quarter notes, eighth notes, and sixteenth notes, with some notes beamed together. There are also rests and repeat signs. The notation is black on a white background.

folkRNN

Compose a tune with a recurrent neural network

PRESS TO GENERATE TUNE

Compose

MODEL

Signature

SEED

718320

METER

MODE

C Major

INITIAL ABC

Write in ABC notation

FOLKRNN

- “recurrent neural network” (RNN), trained using 20k folk songs shared in website <https://thesession.org/> where people upload folk music.
- Copyright uncertain, no copyright statements, no authors listed, no dates.
- Unlikely that any author will come forward to make complaint.



The Drunken Pint by folk-rnn

THE LEGAL SITUATION WITH DATASETS



THE PROBLEM

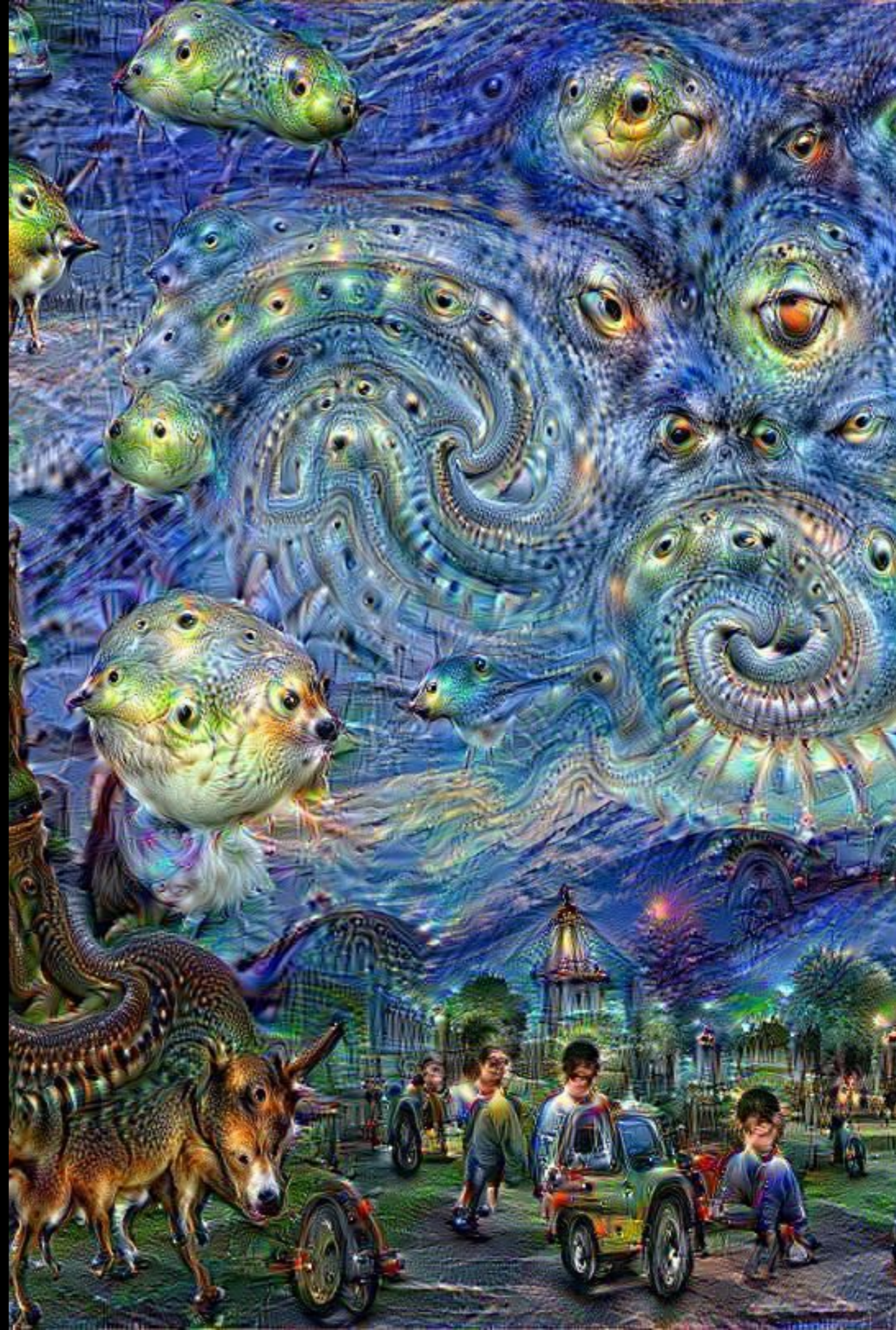


COULD REMBRANDT SUE FOR
COPYRIGHT INFRINGEMENT?



SPOILER ALERT: NO

- Would it ever be possible for an author to sue the maker of a derivative work?
- I don't think so, no requirement for infringement has been met.
- Quite simply, the resulting works do not carry a substantial enough part of the original to be considered infringement.



IS THERE INFRINGEMENT?

- Infringement: to perform one of the exclusive rights without permission, and without a defence.
- Is an AI more akin to a musician that listens to music and is influenced by it?
- Or is it different because copying is involved?

**HOME TAPING IS
KILLING MUSIC**



INFRINGEMENT

- Three elements:
 1. Defendant carries an exclusive act of the owner.
 2. Defendant's work is derived from the copyright work (causal connection).
 3. The work, or a substantial part of the work, has been infringed.



1. RESTRICTED ACTS

- Copy, distribute, perform, lend, rent, adapt, communicated to the public, etc.
- In the case of AI, at most there may have been copying in the “learning” phase.
- Adaptation? Not sure there is a direct connection.



2. CAUSAL CONNECTION

- Francis Day & Hunter v Bron [1963] Ch 587.
- “In a Little Spanish Town” and “Why”. Similar lyrics.
- Claimants must not only prove similarity, but that this similarity was due to an act of copying.
- “...if subconscious copying is to be found, there must be proof (or at least a strong inference) of de facto familiarity with the work alleged to be copied”.

IN A LITTLE SPANISH TOWN

(TWAS ON A NIGHT LIKE THIS)

Lyric by
LEWIS and YOUNG

Music by
MABEL WAYNE

Moderato

Ev-nings are crowd-ed with mem-o-ries, — Thrill-ing me a - gain,
Some-how I keep on re - mem-ber-ing, — Tears that said good - bye,

Like that night in Spain. — Some-one is shar-ing my
Shin-ing in her eye, — Won-der does she keep re-

mem-o-ries — Won-der does she grieve, — Ev-'ry love-ly ev - 'ning:
mem-ber-ing — Prom-is - es of mine — When the moon is shin - ing:

Chords: Bb, F7, Cm7, F7, Bb, Bb7, G7, Cm, G7, Dm7, G7, Cm, C7, F7, F+

© Copyright 1926 Leo Feist, Inc.,
Copyright renewal 1954 Leo Feist, Inc.,
International Copyright Secured
New York, N. Y.
New York, N. Y.
Made in U.S.A.

588R-3

All Rights Reserved Including Public Performance For Profit
Any arrangement or adaptation of this composition without the consent of the owner is an infringement of copyright

3. SUBSTANTIAL USE

- Designer Guild Ltd v Russell Williams (Textiles) Ltd [2001] FSR 113.
- The claimant had created a textile flower design, sued for copyright infringement.
- There was no doubt in the proceedings that the design had copyright and was an original artistic work.
- Deliberate and substantial copying, **“its quality rather than its quantity”**.



INFOPAQ

- “[the process of extraction by Infopaq] increases the likelihood that Infopaq will make reproductions in part within the meaning of Article 2(a) of the [Infosoc] Directive because the **cumulative effect of those extracts** may lead to the **reconstitution of lengthy fragments which are liable to reflect the originality of the work in question**”.

orem ipsum dolor
et, consectetur adipiscing
eiusmod tempor incidunt

incididunt ut labore et do-
a. Ut enim ad minim ve-
ad exercitation ullamco
aliquip ex ea commodo con-
irure dolor in reprehens-
velit esse cillum dolore eu-
ur. Excepteur sint occae-
proident, sunt in culpa
at mollit anim id est labo-
m sit amet, consectetur
d do eiusmod tempor in-
et dolore magna aliqua.
veniam, quis nostrud ex-
o laboris nisi ut aliquip ex
sequat. Duis aute irure

aliquip ex ea commodo
Duis aute irure dolor
voluptate velit esse cillum
nulla pariatur. Excepteur
idatat non proident, sunt
cia deserunt mollit anim
Lorem ipsum sit amet
scing elit, sed do eiusmod
dunt ut labore et dolore
enim ad minim veniam,
citation ullamco laboris
commodo consequat. D
in reprehenderit in volu-
lum dolore eu fugiat nu-
insum dolor sit amet, con-

TEMPLE ISLAND COLLECTIONS V NEW ENGLISH TEAS [2012] EWPC 1



Claimant's work



Defendants' work

TEMPLE ISLAND

- “Although the images undoubtedly differ in their composition, **elements of the overall composition of the claimant’s image have been reproduced.** The **bus** is a Routemaster, driving from right to left with Big Ben on the right of the bus. The **riverside** facade of the Houses of Parliament is part of the image. The bus is on Westminster Bridge (albeit in a different place) in both images. This is obvious in the claimant’s image and can be seen from the presence of the balustrade on the left in the defendants’ image. **There are some people visible** but they are small (and in different places). There is no other obvious traffic. The **edge of Portcullis house** is visible on the right. Running from top to bottom, there is a substantial amount of sky in the picture (albeit more in the claimant’s) and the top of the bus is roughly the same height as the facade of the **Houses of Parliament.**” .

ENGLAND AND WALES
CRICKET BOARD LTD &
ANOR V TIXDAQ LTD &
ANOR [2016] EWHC 575

- “I do not consider that it follows that **reproduction of any part of a broadcast or first fixation amounts to an infringement.** [...] At least in the case of broadcasts and first fixations of films of sporting events, broadcasters and producers invest in the production of broadcasts and first fixations knowing, first, that **some parts of the footage of an event** (e.g. wickets in the case of cricket matches and goals in the case of football matches) **will be more interesting to viewers than other parts** and, secondly, that there is a market for highlights programmes and the like in addition to the market for continuous live coverage.” Arnold J



C-476/17 - PELHAM AND OTHERS V HÜTTER

- Kraftwerk's "Metall auf Metall".
- "the phonogram producer's exclusive right under that provision to reproduce and distribute his or her phonogram allows him to prevent another person from taking a sound sample, even if very short, of his or her phonogram for the purposes of including that sample in another phonogram, unless that sample is included in the phonogram in a modified form unrecognisable to the ear."
- "a phonogram which contains sound samples transferred from another phonogram does not constitute a 'copy', within the meaning of that provision, of that phonogram, since it does not reproduce all or a substantial part of that phonogram."



DERIVATIVE USES

- Non-harmonised area of law, exceptions and limitations are all over the place.
- Fair use - Transformative use
- Fair dealing
- Adaptation
- Derivative work



CONCLUDING



"We seem to be made to suffer. It's our lot in life"

C3PO

@TECHNOLLAMA

