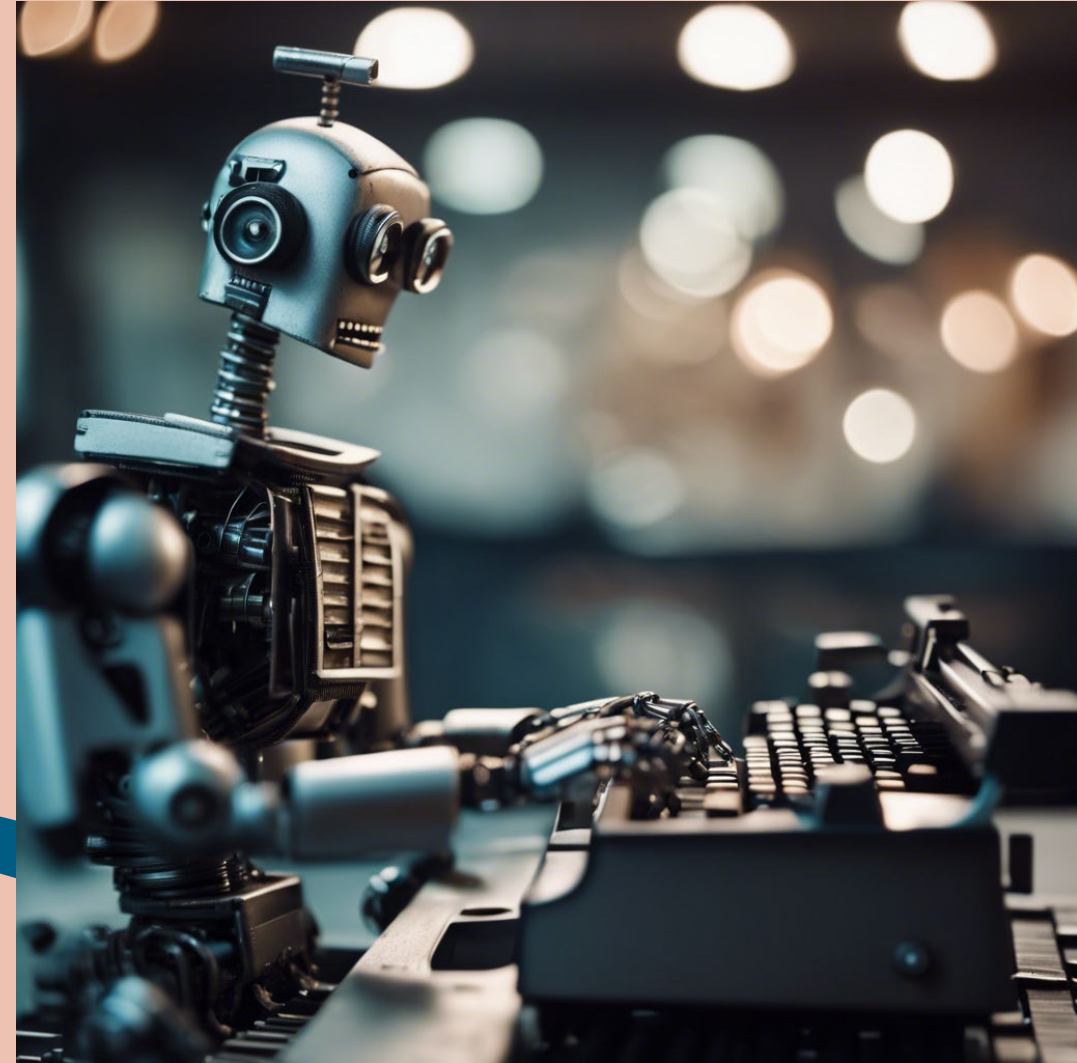


Bird & Bird

Openli Webinar AI and Copyright

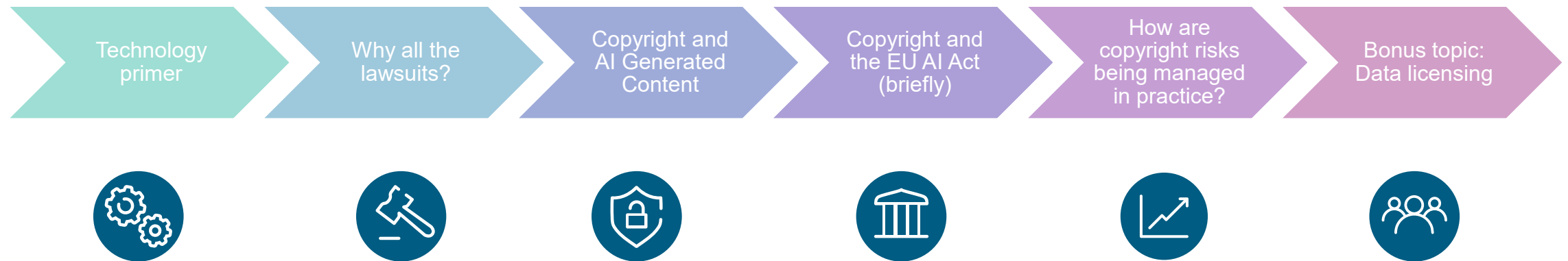
Toby Bond, Partner, London
(toby.bond@twobirds.com)

17 April 2024



Created by AI

Agenda

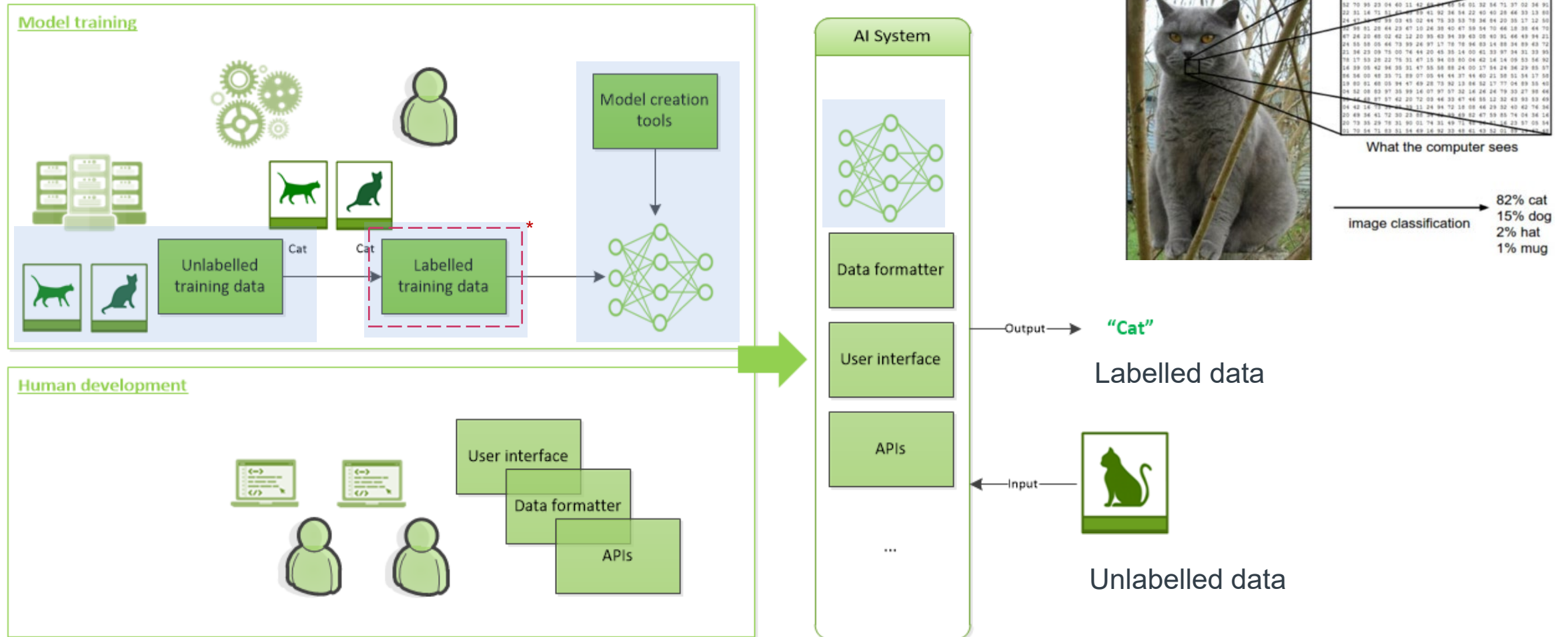


01

Technology Primer



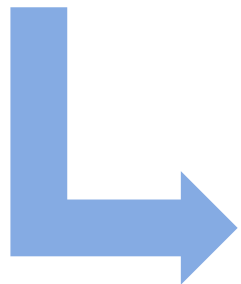
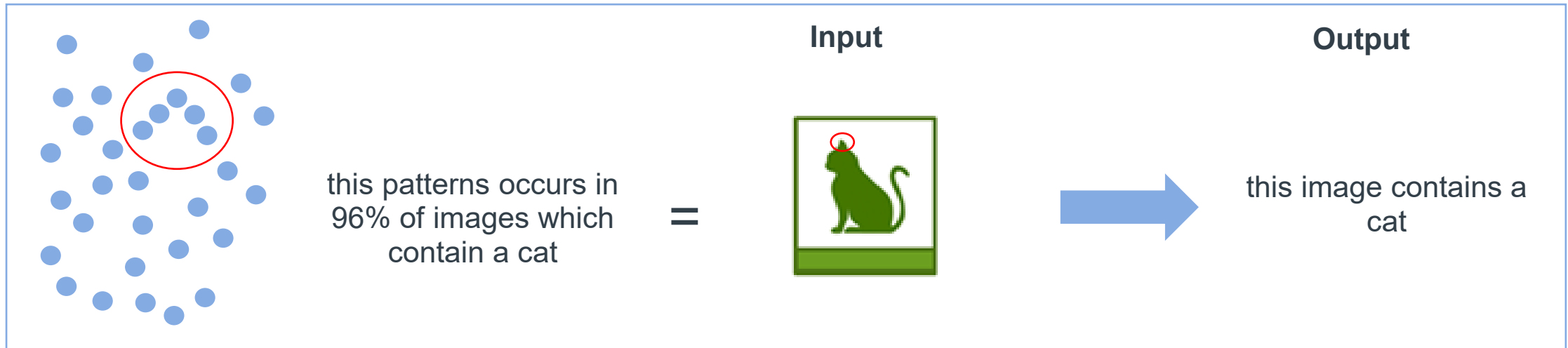
Predictive AI: Machine Learning



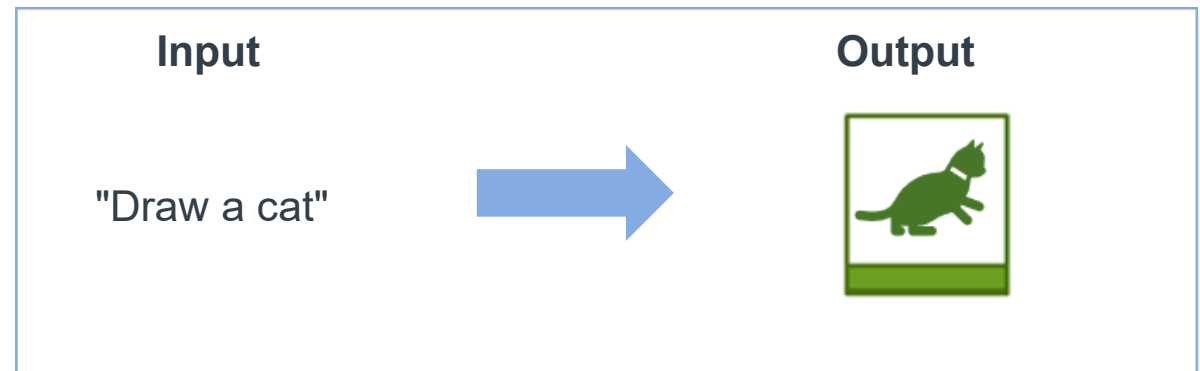
* labelled training data is only required for some types of machine learning

What is Generative AI?

Machine Learning: finding patterns in data



Generative AI: making new examples of patterns



What is Generative AI?

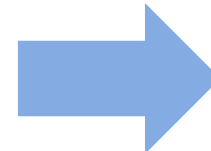
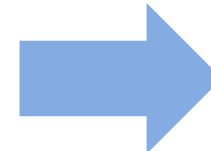
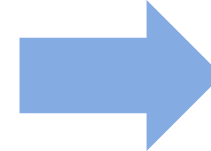
Examples of image, text and code creation

Input

"a robot artist with an easel painting a landscape of rolling hills" (style: comic book)

write a haiku about a robot artist with an easel painting a landscape of rolling hills

write html code to embed the hakiu in a website



Output



*Metal hands create
A vision of nature's grace
On a canvas white*

```
<html>
<head>
  <style>
    p {
      font-family: Arial, sans-serif;
      font-size: 16px;
      text-align: center;
    }
  </style>
</head>
<body>
  <p>Metal hands create<br>
  A vision of nature's grace<br>
  On a canvas white</p>
</body>
</html>
```

Some terms we'll be using to explore the legal issues

INPUTS / PROMPTS

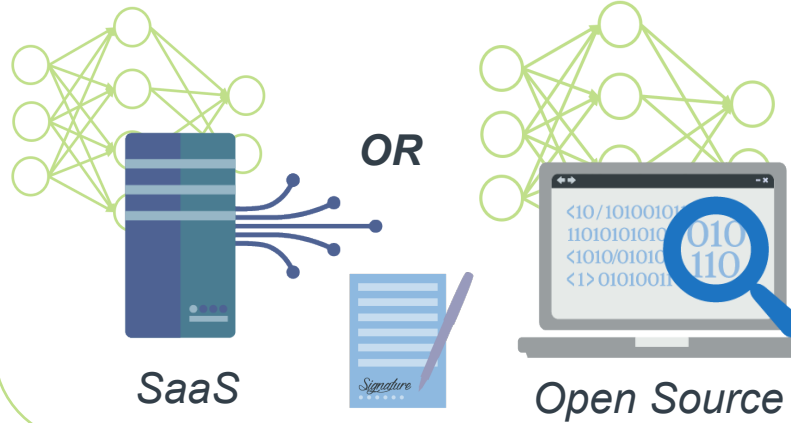
*write a haiku about a robot artist with
an easel painting a landscape of
rolling hills*

*write html code to embed the hakiu on
a website*

*write html code to embed
the hakiu on a website*



AI TOOLS



OUTPUTS



Created by AI

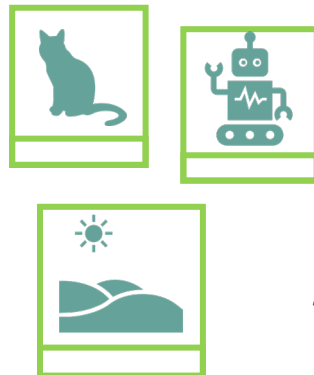
*Metal hands create
A vision of nature's grace
On a canvas white*



Created by AI

```
<html>
<head>
  <style>
    p {
      font-family: Arial, sans-serif;
      font-size: 16px;
      text-align: center;
    }
  </style>
</head>
<body>
  <p>Metal hands create<br>
  A vision of nature's grace<br>
  On a canvas white</p>
</body>
</html>
```

TRAINING DATA



MODELS



Training process

It's actually a
bit more
complicated ...
but we'll come
back to this
later

02

Why all the lawsuits?



There have been quite a few...

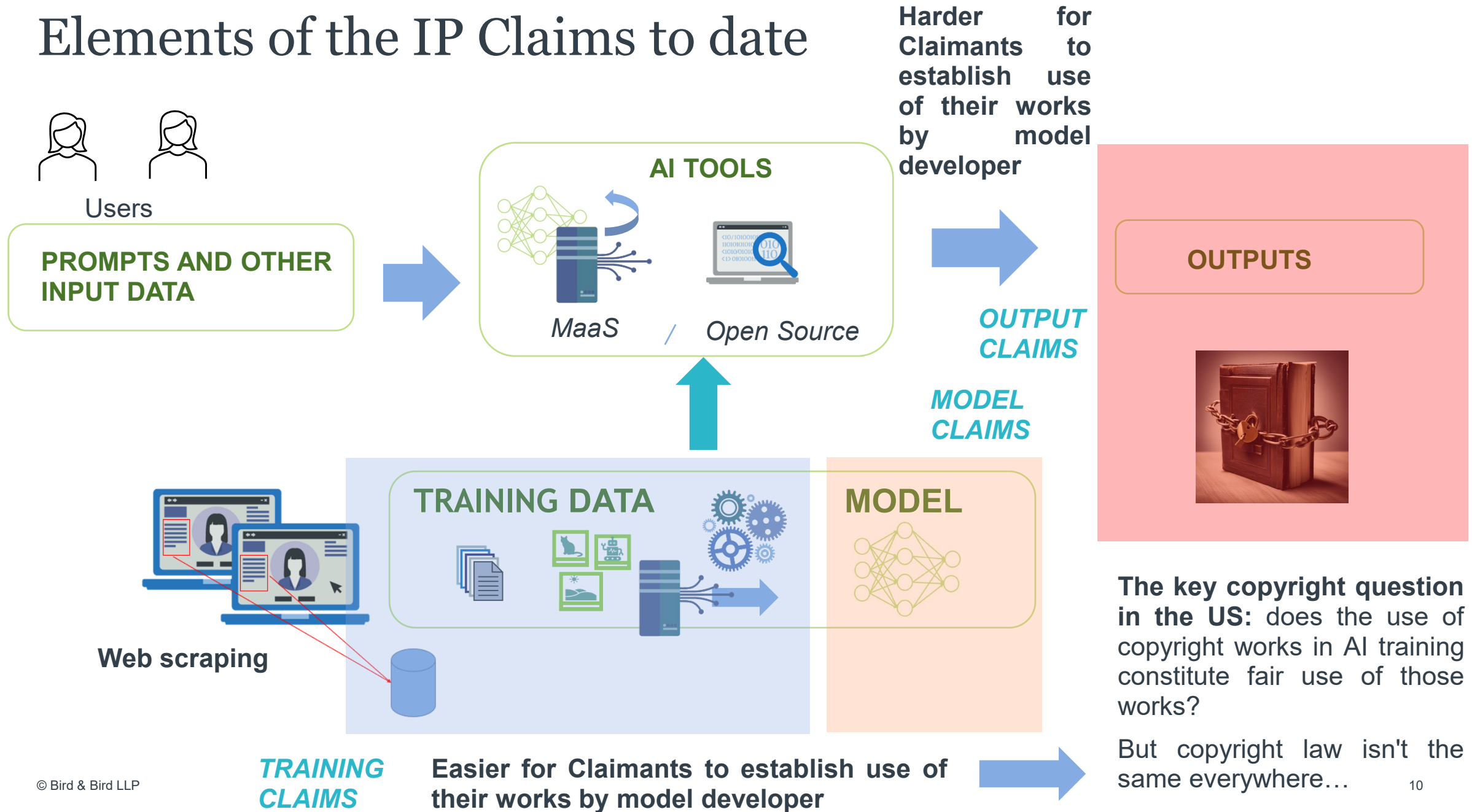


13	Month	Parties	Copyright works	Product(s)	Venue
1	Nov 22	Doe v GitHub, OpenAI, Microsoft	Code	Github CoPilot	N.D. Cal, USA
2	Jan 23	Anderson v Stability, Midjourney, DeviantArt	Images	Stable Diffusion, Midjourney, Deviant Art	N.D. Cal, USA
3	Jan 23	Getty v Stability	Images	Stable Diffusion	High Court, UK
4	Feb 23	Getty v Stability	Images	Stable Diffusion	D. Del, USA
5	Jun 23	Tremblay v Open AI	Text/Books	GPT	N.D. Cal, USA
6	July 23	Silverman v Meta	Text/Books	LLaMA	N.D. Cal, USA
7	July 23	Silverman v OpenAI	Text/Books	GPT	N.D. Cal, USA
8	July 23	J.L v Alphabet, Google, DeepMind	Text/Books, Images, music, video	PaLM-2, Imagen, MusicLM, DuetAI, Gemini	N.D. Cal, USA
9	Sep 23	Chanbon v Open AI	Text/Books	GPT	N.D. Cal, US
10	Sep 23	Authors Guild v Open AI	Text/Books	GPT	S.D. NY, US
11	Oct 23	Huckabee v Meta, Bloomberg, Microsoft	Text/Books	LLaMa, Bloomberg GPT	S.D. NY, US
12	Oct 23	UMG v Anthropic	Text/Song lyrics	Claude	M.D. Tennessee USA
13	Nov 23	Sanction v OpenAI, Microsoft	Text/Books	GPT	S.D. NY, US
14	Dec 23	NTY v OpenAI, Microsoft	Text/News content	GPT	S.D. NY, USA
15	Feb 24	Shanghai Character Licence Administrative Co v "Tab"	Images/Ultraman	Undisclosed image generator	Guangzhou Internet Court

Three proposed US class actions also relate to privacy issues

Four more US copyright claims filed in 2024 (so far)...

Elements of the IP Claims to date



AI Training Data

Public websites

Copyright/database rights and exceptions

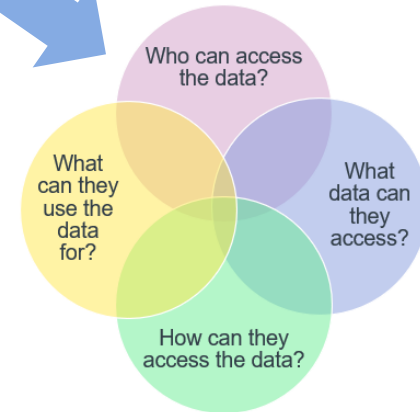
- Is the training data protected by copyright or database rights?
- Do any exceptions apply which cover obtaining the data and using the data for AI training?
- Where is the training taking place, and which exceptions apply?

Other web scraping issues

- Is there a breach of website T&Cs?
- Do any tortious claims apply?
- Are anti-hacking laws being breached?

Other contracts

- What are the restrictions on accessing, using and disseminating data?
- Key terms: confidentiality, IP licence, data licence
- Also interplay with data protection terms (processors can only do what they're told with personal data).



Open source/creative commons licence terms

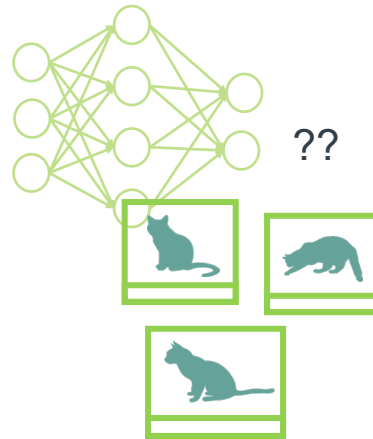
- Do the OS terms permit use for AI training (for a commercial purpose)?
- Are there attribution requirements/provisions on derivative works?

Can be complex questions due to range of legal rights involved and lack of harmonisation

AI Models and Outputs

Training data risks for models

- Is a trained model a "derivative work" of the training data under US copyright law (such that using, offering, disseminating the model infringes any copyright in the training data)?
- Is a trained model an "infringing copy" under UK copyright law and does bringing it into the UK result in secondary infringement?
- Is a trained model a "derivative work" under any open source/creative commons licence which applied to the training data?



Really significant question for generative AI

Hard questions at the cutting edge of law and technology

Training data risks for outputs

- Do any AI outputs reproduce all or a substantial part of a copyright work used as an input?
- If they do, is the user responsible due to the prompts they selected?
- Is an AI output a "derivative work" under any open source/creative commons licence which applied to (i) the training data; or (ii) the model?



Hard questions again

Key Takeaways: Litigation risk



- All the litigation to date (except Ultraman) has **targeted generative AI model developers**.
- However, claims relating to outputs (and possibly models) **could potentially be brought against users**.
- There are lots of **unresolved issues** which are being addressed in these litigations:
 - does US fair use apply to generative AI development?
 - can models themselves be an infringement of copyright?
 - can outputs infringe, and in what circumstances?
 - if outputs can infringe, how is responsibility for outputs allocated between model developers and users?
- Some providers are offering **IP indemnities** to reassure users (we'll come back to this later).
- The **distinction between developers, deployers and users can get blurry** (we'll come back to this later as well).

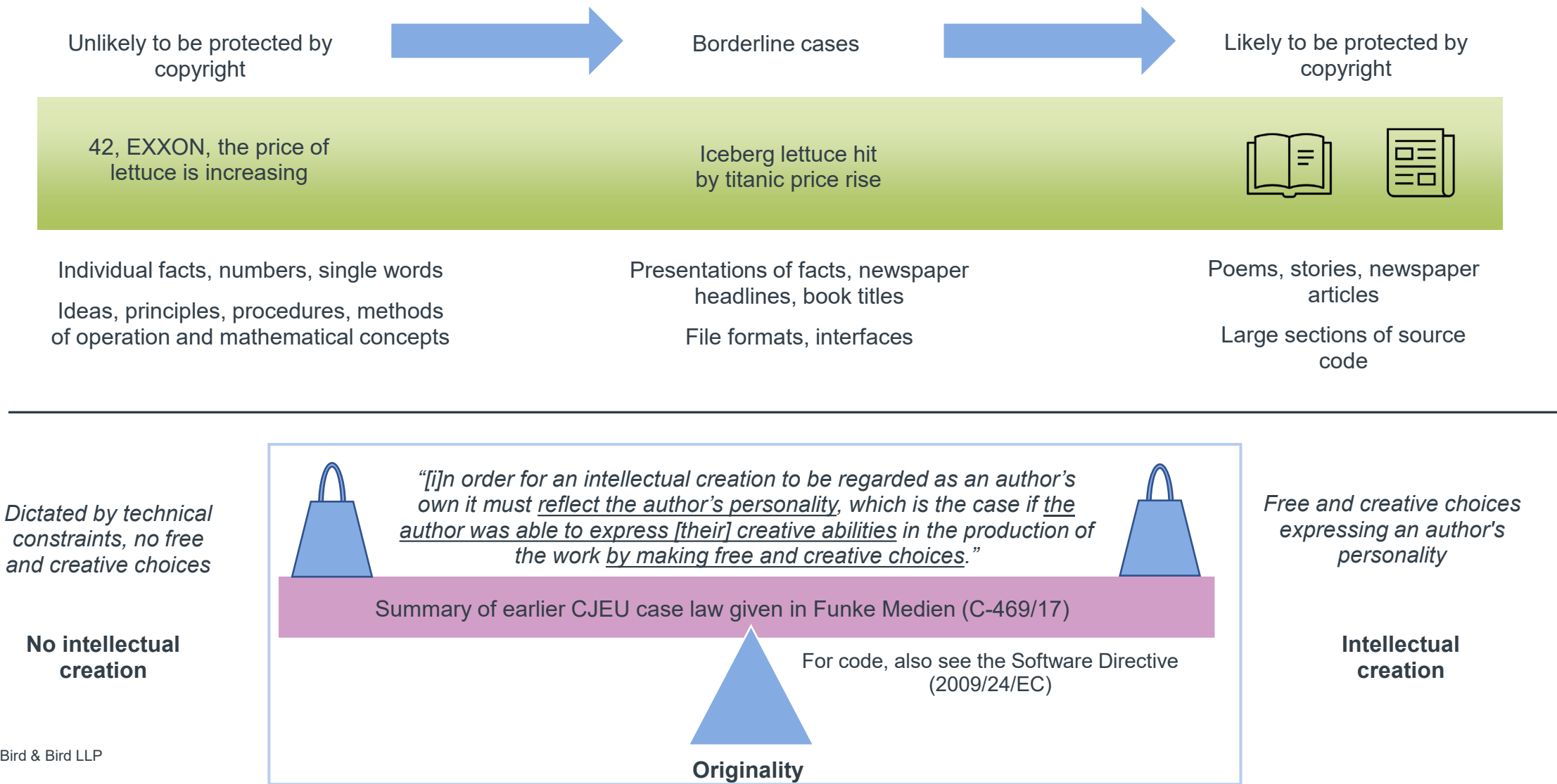
03

Copyright and AI Generated Content

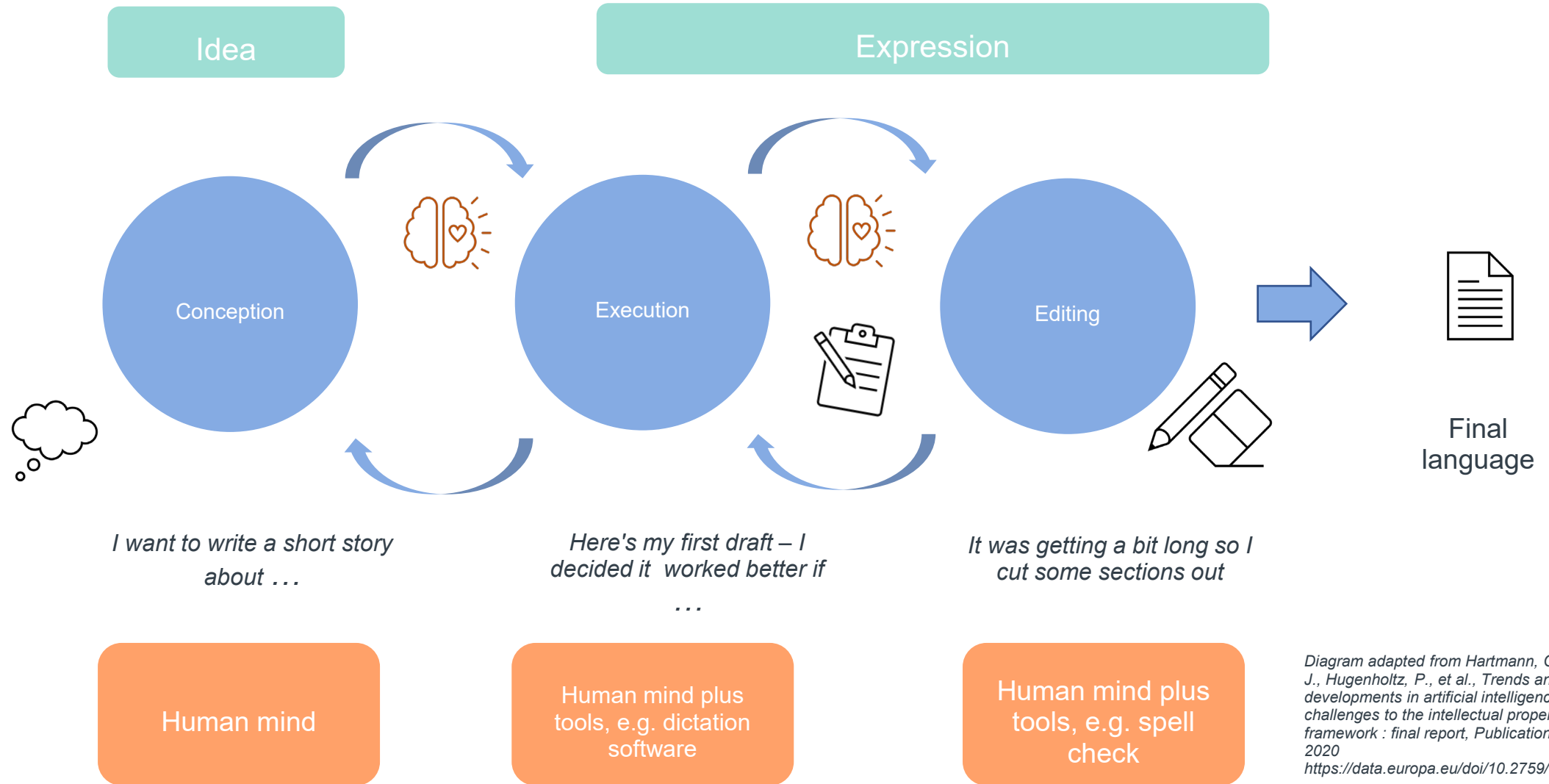


Copyright protection for AI Outputs:

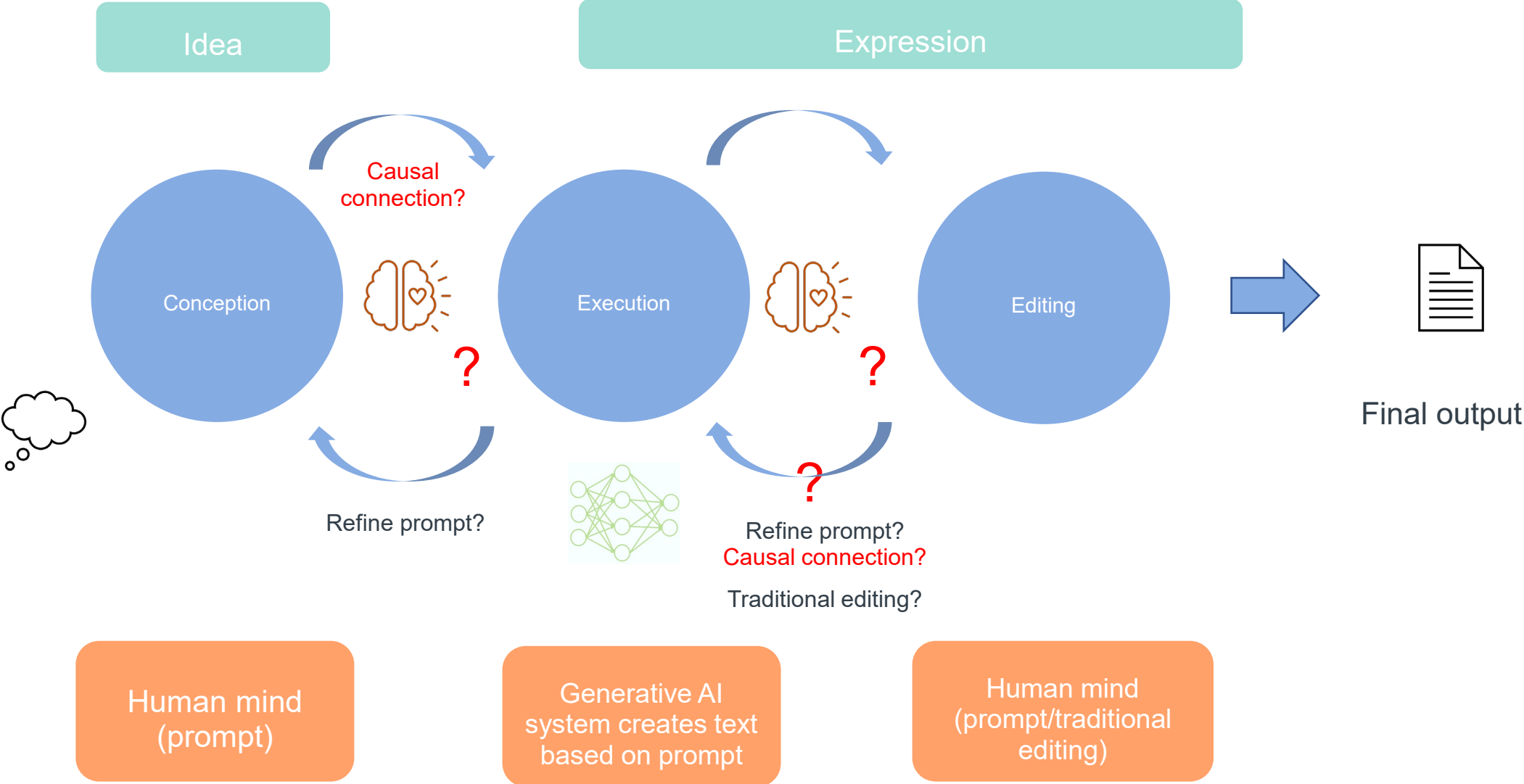
The European perspective: When does copyright protect language and code?



The creation process: without Generative AI



The creation process: with Generative AI



US Copyright Office Guidelines (March 2023)

Principles

- Copyright can protect only material that is the product of human creativity, i.e. there is a human authorship requirement for registration. The Office will begin by asking whether the 'work' is:
- (1) *basically one of human authorship, with the computer [or other device] merely being an assisting instrument, or*
- (2) *whether the traditional elements of authorship in the work (literary, artistic, or musical expression or elements of selection, arrangement, etc.) were actually conceived and executed not by man but by a machine."*

Application to AI-generated material

- For work containing AI-generated material are the AI contributions (i) the result of mechanical reproduction or; (ii) an author's own original mental conception, to which [the author] gave visible form? When an AI technology determines the expressive elements of its output, the generated material is not the product of human authorship.
- When an AI technology receives solely a prompt from a human and produces complex written, visual, or musical works in response, the "traditional elements of authorship" are determined and executed by the technology—not the human user. The prompts function more like instructions to a commissioned artist.
- A human may select or arrange AI-generated material in a sufficiently creative way that "the resulting work as a whole constitutes an original work of authorship."
- Or an artist may modify material originally generated by AI technology to such a degree that the modifications meet the standard for copyright protection. In these cases, copyright will only protect the human-authored aspects of the work, which are "independent of" and do "not affect" the copyright status of the AI-generated material itself.

Registration guidance

- Individuals who use AI technology in creating a work may claim copyright protection for their own contributions to that work.
- AI-generated content that is more than *de minimis* should be explicitly excluded from the application.

Standard Application

1. **"Author Created" field:** describe the authorship that was contributed by a human:

- **Example 1:** *[describe portions of the textual work that are human-authored]*
- **Example 2:** Selection, coordination, and arrangement of *[describe human-authored content]* created by the author and *[describe AI content]* generated by artificial intelligence

2. **"Limitation of the Claim" field (Other/Material Excluded):** brief description of the AI-generated content.

Alternative: *Applicants who are unsure of how to fill out the application may simply provide a general statement that a work contains AI-generated material. The Office will contact the applicant when the claim is reviewed and determine how to proceed.*



LIBRARY OF CONGRESS
Copyright Office
37 CFR Part 202
Copyright Registration Guidance:
Works Containing Material Generated
by Artificial Intelligence
AGENCY: U.S. Copyright Office, Library
of Congress.
ACTION: Statement of policy.
SUMMARY: The Copyright Office issues
this statement of policy to clarify its
practices for examining and registering
works that contain material generated
by the use of artificial intelligence
technology.

What about the UK provisions on computer-generated works (and similar provisions in Ireland, India, New Zealand, South Africa)?

Copyright, Designs and Patents Act 1988

s1(1) Copyright is a property right which subsists in accordance with this Part in the following descriptions of work— (a) original literary, dramatic, musical or artistic works, ...

s9(3) In the case of a literary, dramatic, musical or artistic work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken.

s178 “computer-generated”, in relation to a work, means that the work is generated by computer in circumstances such that there is no human author of the work;

Term of protection: 50 years from creation

UKIPO Consultation (October 2021)

From a legal perspective, a computer-generated work must be original if it is to receive protection. But the legal concept of originality is defined with reference to human authors and characteristics like personality, judgement and skill. It has been argued that the law is unclear and contradictory.

Proposals for consultation

Option 0: Make no legal change

Option 1: Remove protection for computer-generated works

Option 2: Replace the current protection with a new right of reduced scope/duration

UK Government response to consultation (June 2022)

21. We have decided to adopt Option 0: make no changes to the law. There is no evidence at present that protection for CGWs is harmful, and the use of AI is still in its early stages. As such, a proper evaluation of the options is not possible, and any changes could have unintended consequences. But we will keep the law under review and could amend, replace or remove protection in future if the evidence supports it.

Did we decide whether the law is unclear and contradictory?

But don't forget about related rights...

Films and sound recordings (which are not copied from an earlier film or sound recording) are protected as copyright works in the UK without an originality requirement, with the author being the principal director and producer (for a film) and the producer (for a sound recording). Films and sound recordings captured by AI will therefore be protected by copyright* in the UK.

*This is not copyright in the Berne Convention sense and is really a related economic right.

Key Takeaways: Copyright Subsistence



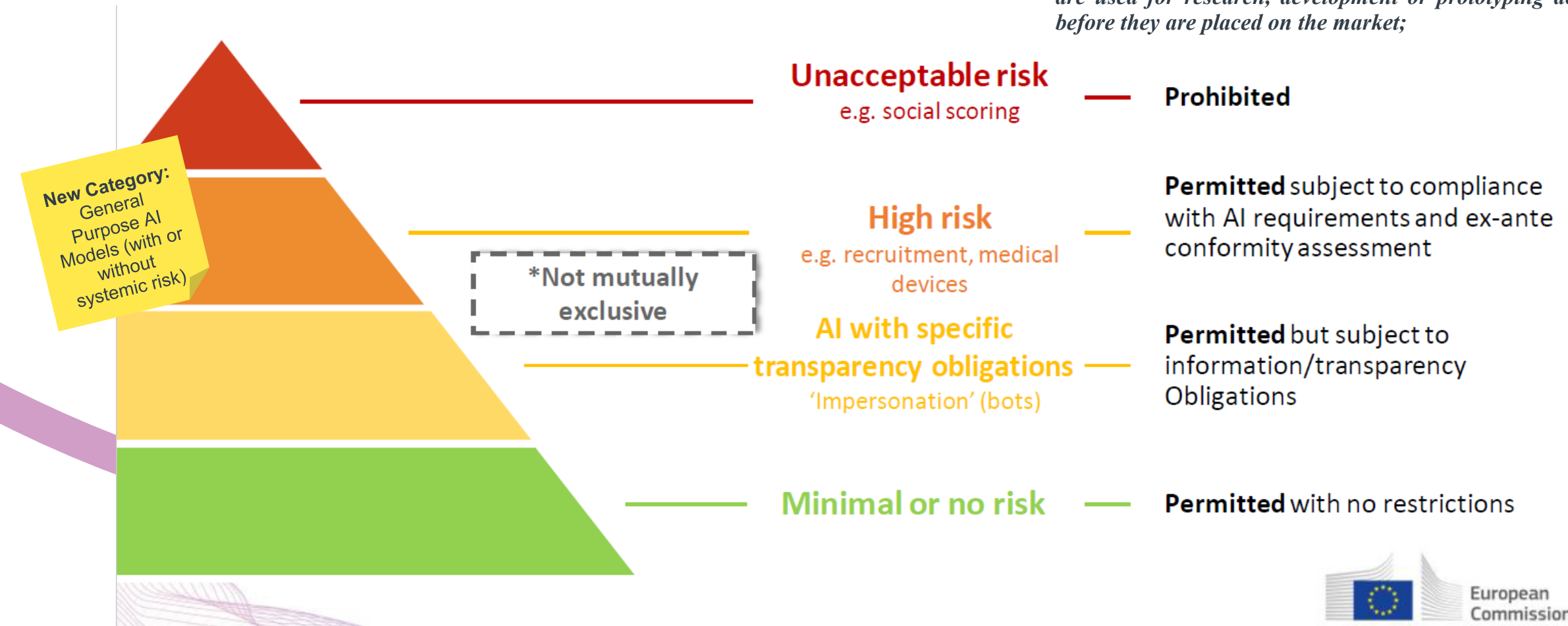
- **A general "rule of thumb":** outputs generated by a simple text prompt typically won't be protected by copyright.
- **BUT, beyond that simple scenario it all depends on:**
 - How is the content being generated – does it allow room for some human creativity in the output?
 - Is the content being edited / manipulated / incorporated into other content after generation?
 - Which jurisdiction do we want to protect the output in – not every jurisdiction is the same.
- **The key commercial question:** is it important that we get copyright protection for this / can we protect it in some other way?

04

Copyright and the EU AI Act (briefly)

The EU AI Act takes a risk-based approach

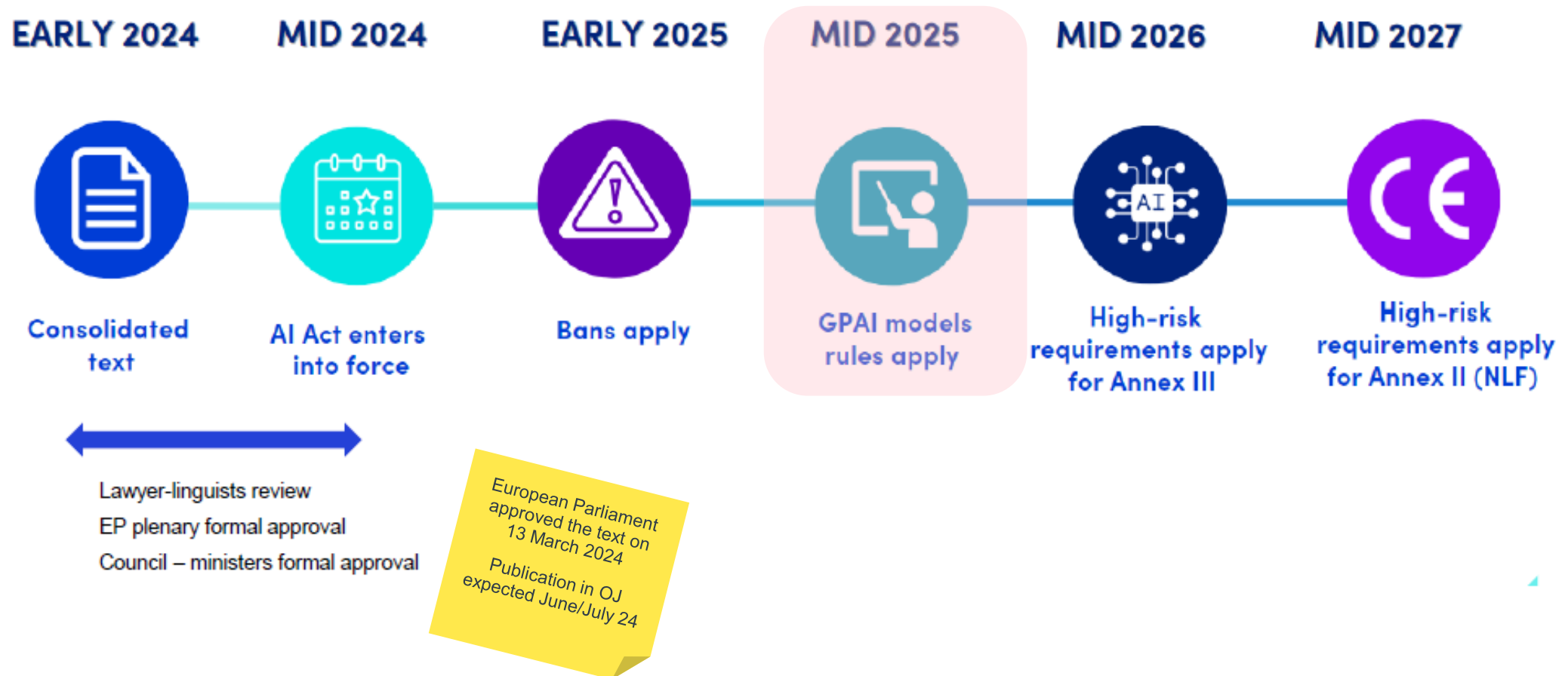
*'general-purpose AI model' means an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, **that displays significant generality and is capable of competently performing a wide range of distinct tasks** regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market;*



Copyright obligations for providers of general-purpose AI systems

- Providers of a general-purpose AI model must:
 - have "*a policy to comply with Union copyright law*", including complying with opt-outs from the EU's commercial TDM exception using "state of the art technologies" - Recital (105) and Article 53(1)(c).
 - disclose details of the content used for training - Recital (107) and Article 53(1)(d).
 - "*for example by listing the main data collections or sets that went into training the model, such as large private or public databases or data archives, and by providing a narrative explanation about other data sources used*"
- The AI Office will prepare a code of practice explaining the level of detail required for training data disclosures (expected Q2 2025).
- No exception for open source general-purpose AI models - Recital (104).
- Providers fine-tuning general-purpose AI models need to disclose the new training data sources - Recital (109).

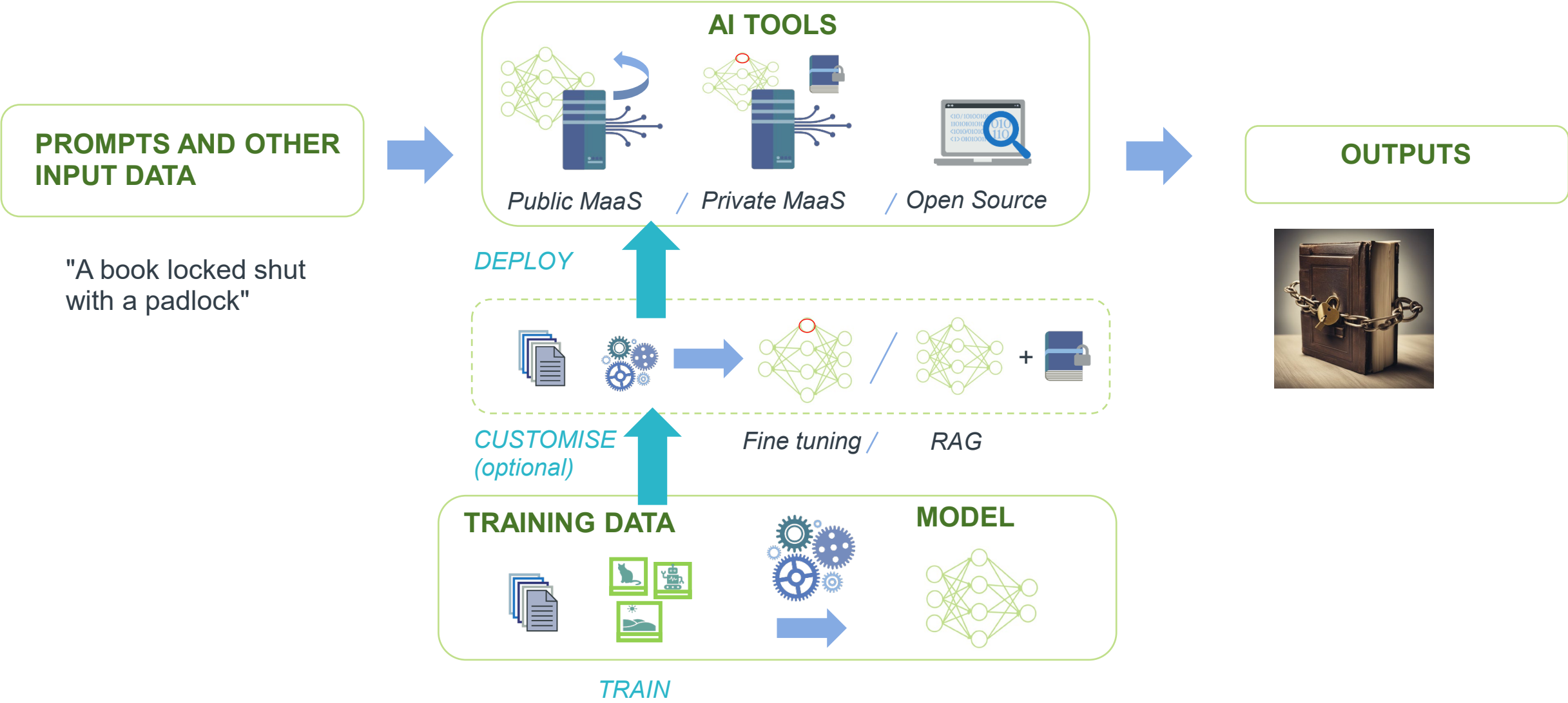
Timeline



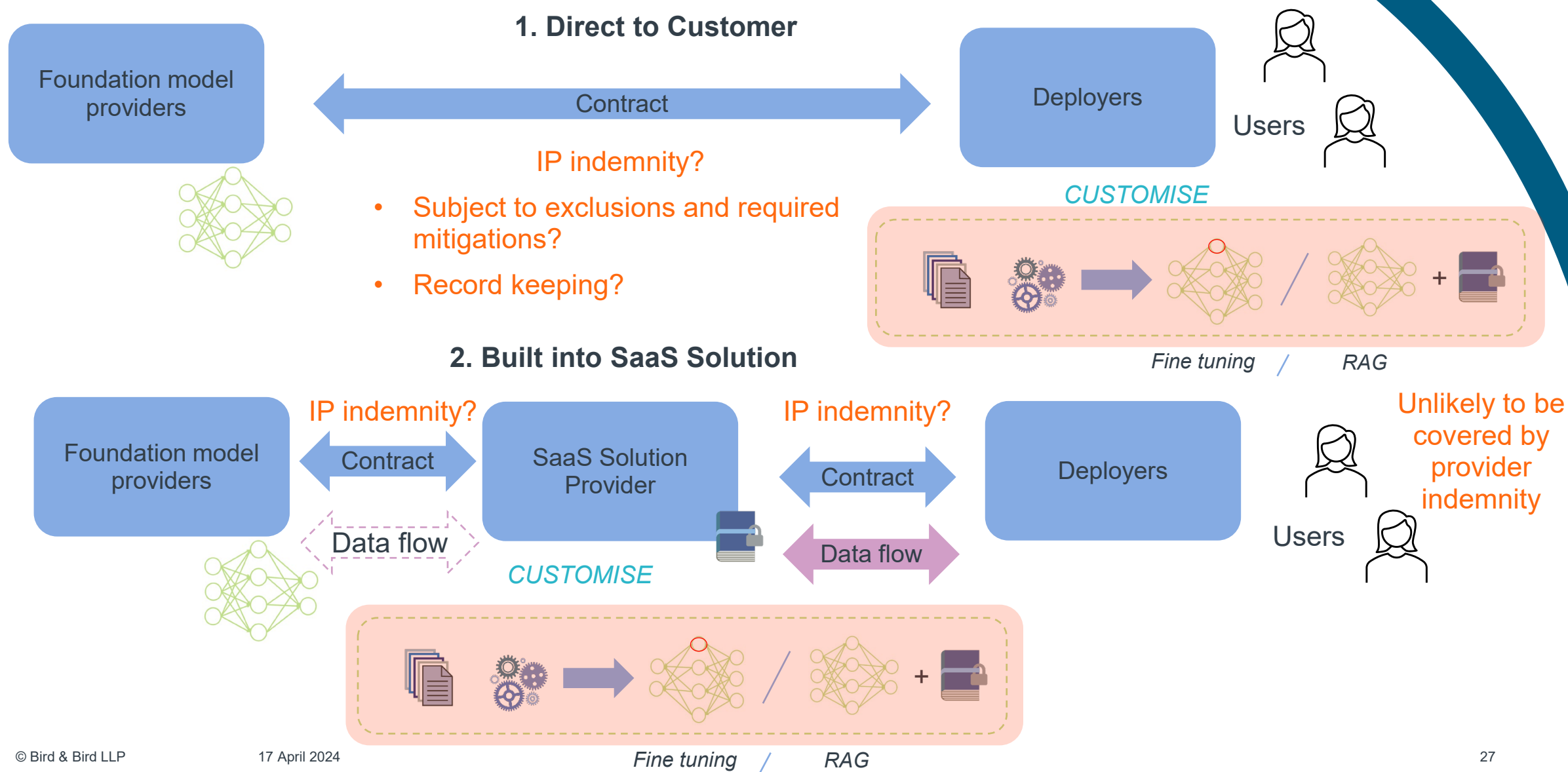
05

How are copyright risks
being managed in practice?

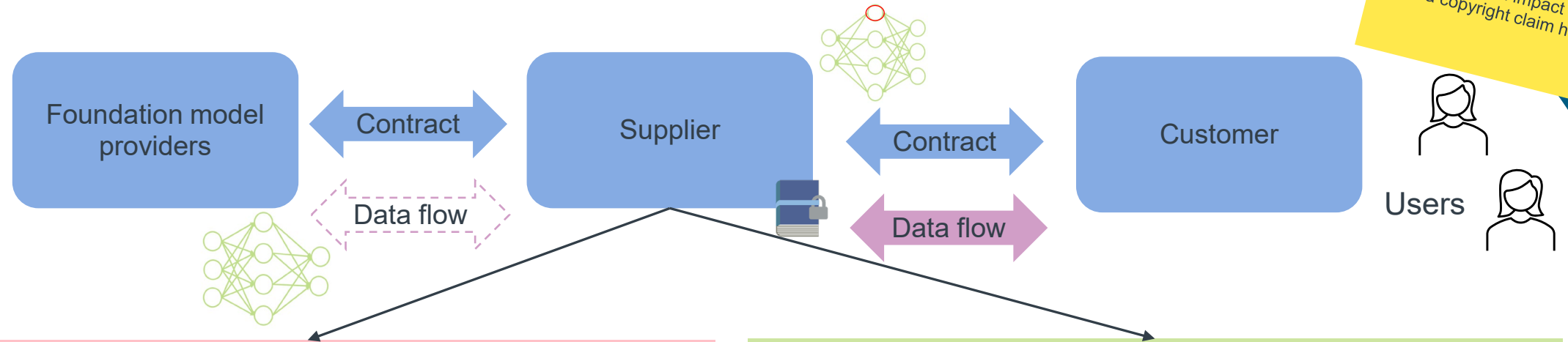
Elements of Generative AI Deployments



IP Risk in the AI value chain



Supply chain examples



Example 1: Creative agency

Lower risk

- AI only used for ideation, AI tools used to modify or enhance content we're permitted to use, AI elements of deliverables will be very small/limited

Higher risk

- Key elements of a campaign will be AI generated, high profile campaigns

Example 2: Outsourced software development

Lower risk

- AI only used for coding suggestions / autocomplete, code won't be shared outside of the organisation (e.g. SaaS platform), code could be easily removed / replaced

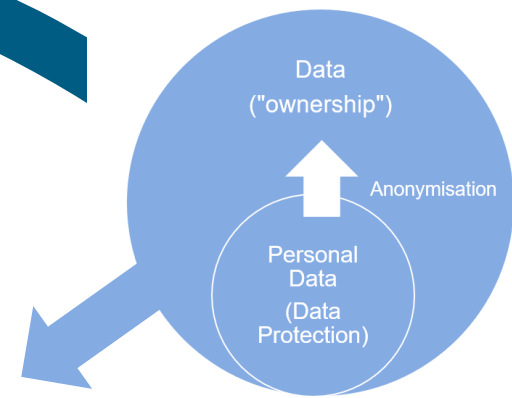
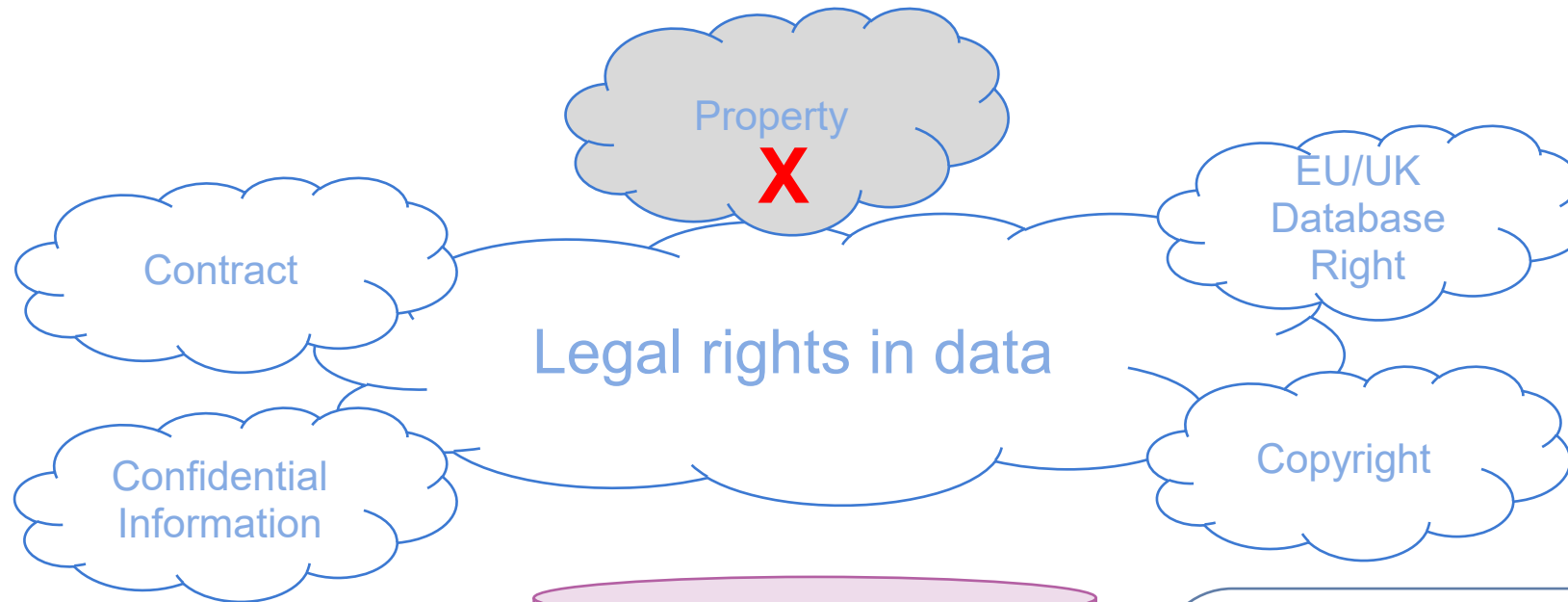
Higher risk

- Code will be committed to public repositories, licences / trade secrets protection won't be enough

06

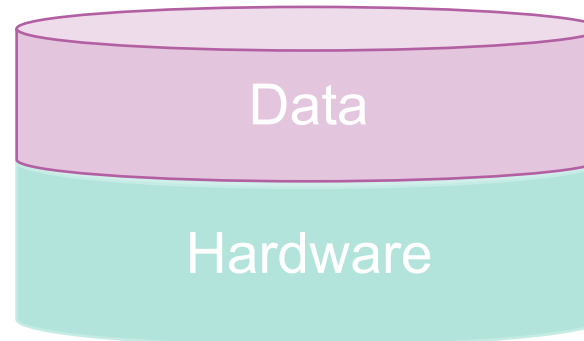
Bonus Topic: Data Licensing

Data "ownership"

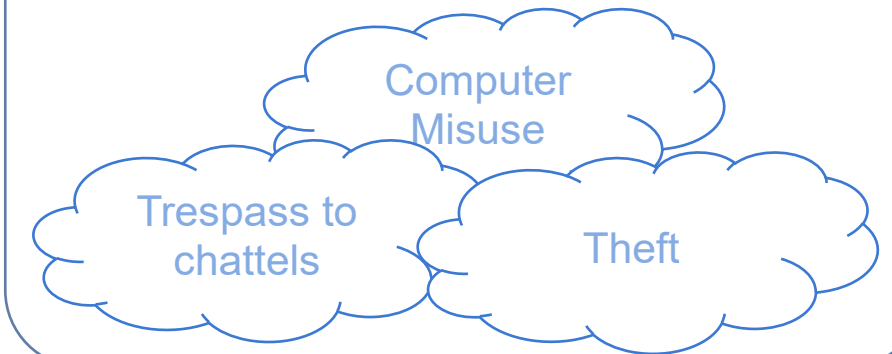


Related area

Competition law, e.g. Articles 101 and 102 TFEU

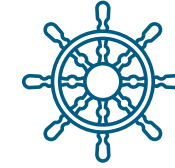


Property laws which can be relevant to data

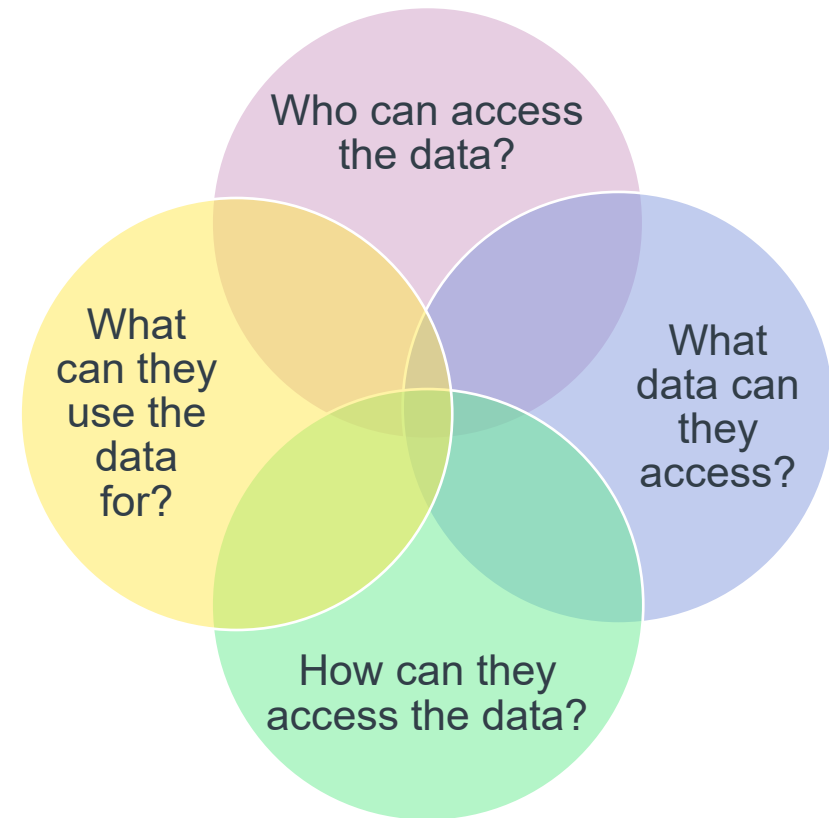


Where to start with a Data Licence

Consider Four Dimensions of Control

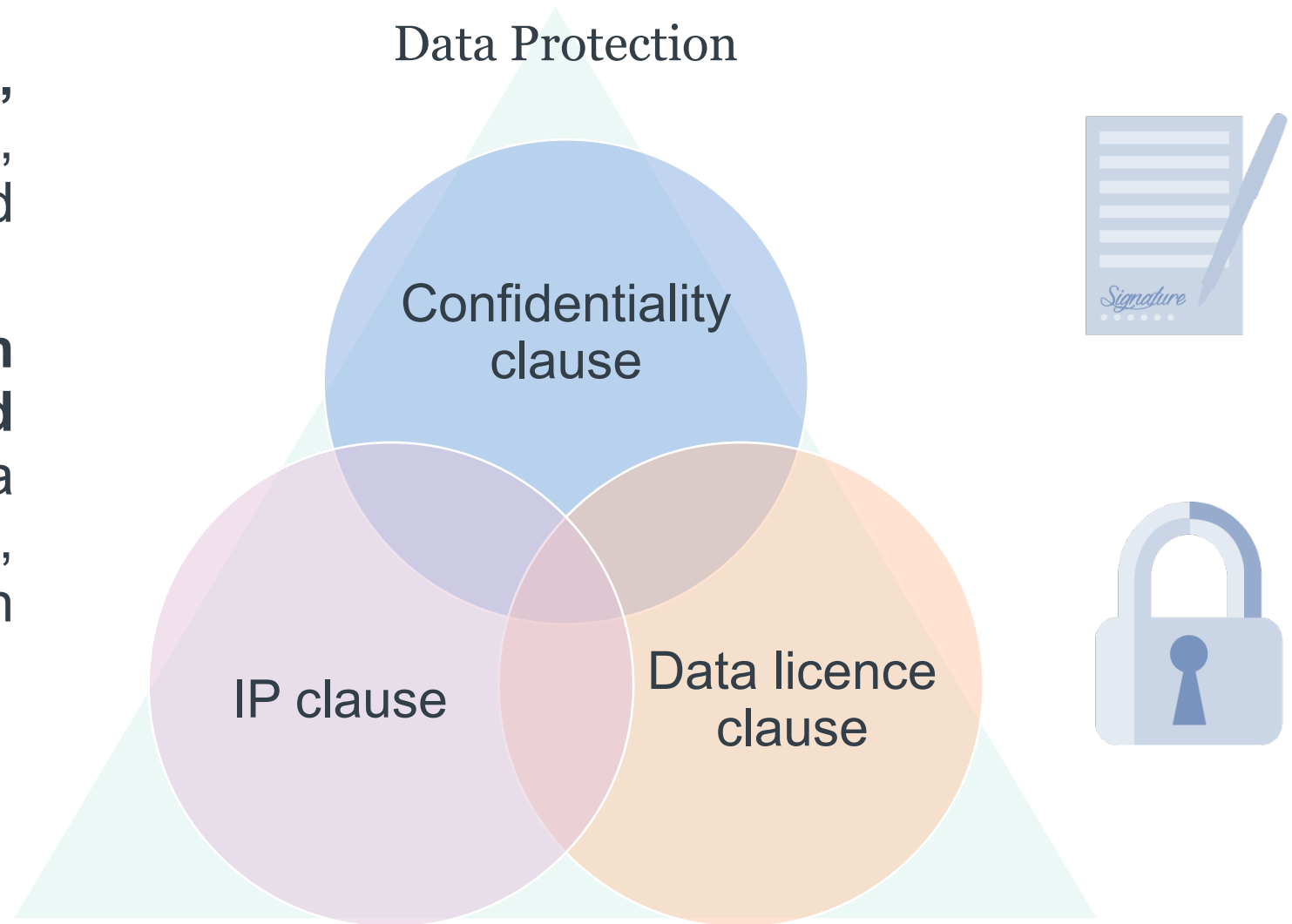


1. Who can access the data?
2. What data can they access?
3. How can they access the data?
4. What can they use the data for?



Where to implement decisions

- **Frame key definitions**, e.g. data, derived data, metadata, IP and confidential information.
- **Align clauses which deal with access and use of data**, e.g. data licence, IP licence, confidential information clauses.



Key Takeaways: Data Licensing



- Data may or may not be protected by IP rights - data licences usually draft round this.
- Consider the four dimensions of control when drafting a data licence and implement those decisions in the relevant clauses and definitions.
- Work through the licence scope using IP licence principles – then apply data licence specific points.
- You'll need to craft your own definition of derived data – which now needs to think about use as AI training data.
- Warranties and audit clauses are other key commercial terms.
- Termination provisions need careful attention – consider ongoing licences and deletion obligations for each category of data.

Thanks for listening...
... any questions?





Thank you

twobirds.com

Abu Dhabi • Amsterdam • Beijing • Bratislava • Brussels • Budapest • Casablanca • Copenhagen • Dubai • Dublin • Dusseldorf
• Frankfurt • The Hague • Hamburg • Helsinki • Hong Kong • London • Lyon • Madrid • Milan • Munich • Paris
• Prague • Rome • San Francisco • Shanghai • Shenzhen • Singapore • Stockholm • Sydney • Warsaw

The information given in this document concerning technical legal or professional subject matter is for guidance only and does not constitute legal or professional advice. Always consult a suitably qualified lawyer on any specific legal problem or matter. Bird & Bird assumes no responsibility for such information contained in this document and disclaims all liability in respect of such information.

This document is confidential. Bird & Bird is, unless otherwise stated, the owner of copyright of this document and its contents. No part of this document may be published, distributed, extracted, re-utilised, or reproduced in any material form.

Bird & Bird is an international legal practice comprising Bird & Bird LLP and its affiliated and associated businesses.

Bird & Bird LLP is a limited liability partnership, registered in England and Wales with registered number OC340318 and is authorised and regulated by the Solicitors Regulation Authority (SRA) with SRA ID497264. Its registered office and principal place of business is at 12 New Fetter Lane, London EC4A 1JP. A list of members of Bird & Bird LLP and of any non-members who are designated as partners, and of their respective professional qualifications, is open to inspection at that address.