Intangible Cultural Heritage, AI and data governance

What does data governance for ICH looks like in the age of AI?

2021 — UNESCO Recommendation on the Ethics of AI

2022

High- level principles on Ethical AI proliferate

2023 — There is an international consensus: 'Ethical' or 'Responsible' AI is the likely solution to access the benefits of AI, limit the risks or downsides.

2024 — What constitutes 'Ethical' or 'Responsible' AI for culture generally, and for Intangible Cultural Heritage specifically is unclear.

So, where shall we start?

2025

Data, and data governance
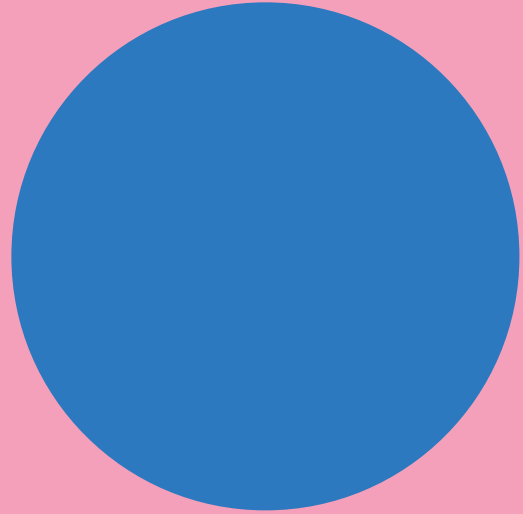
"Data governance" or ICH data governance

- We refer to **data governance** as
   The principles, habits, tools and processes we can put in place to make sure machine-readable content (or 'data') is used, shared and protected in a manner consistent with chosen values (like consent, control, compensation, authenticity or privacy)

- In many ways, the heritage sector has done data governance for decades, when you do 'provenance' work for example, or when you do rights management related to heritage materials.

- Similarly, communities have also managed data governance, and those principles and practices are being transposed into digital environments, including AI. For example, the FAIR and CARE principles are principles and tools of data governance.
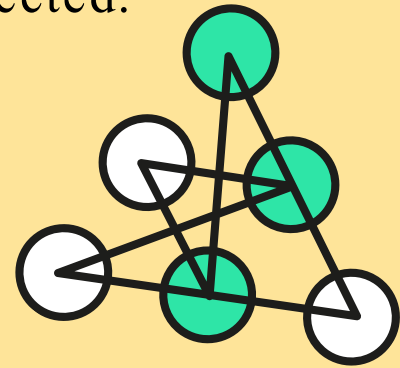
Objective

Give you a practical starting point to think about how ICH data flows through the AI lifecycle to inform decisions about how to share ICH, engage with ICH or use AI in your ICH practice.
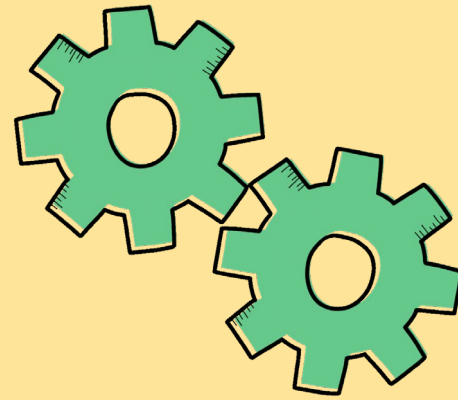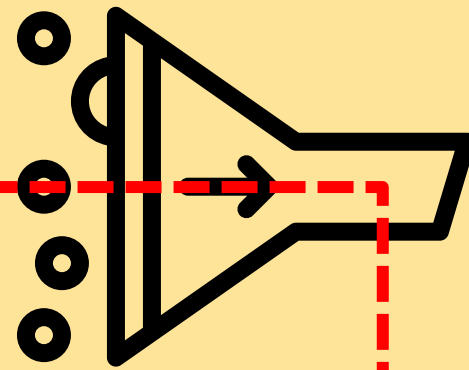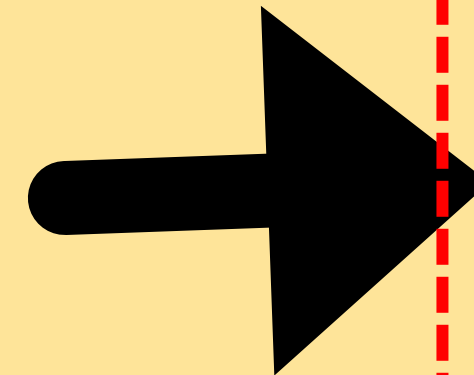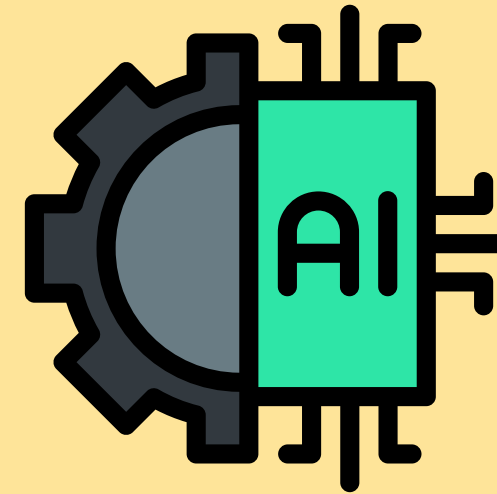
How an AI model operates

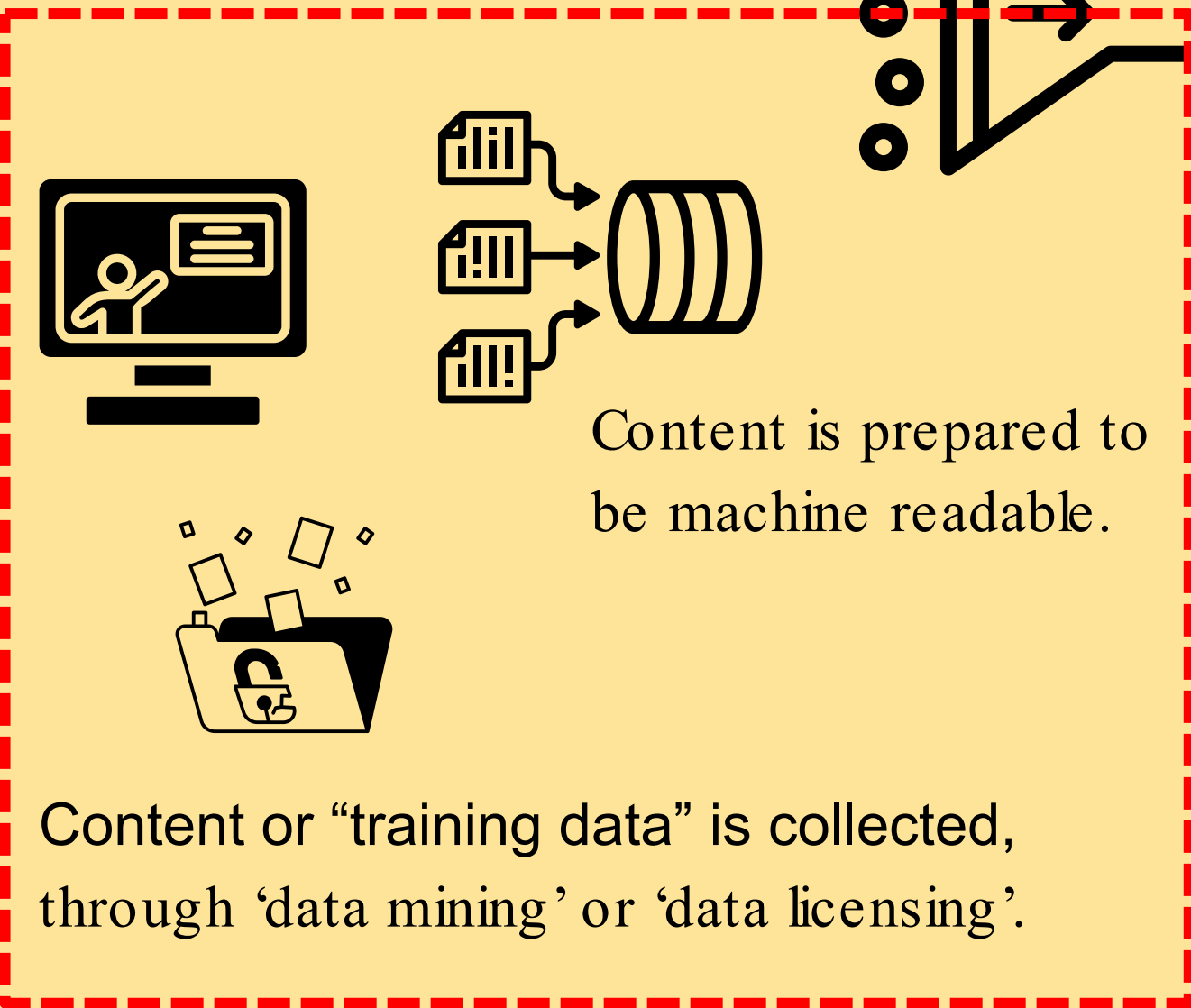The right mathematical formula (or 'algorithm') for the task is selected.

Training data is ingested by the algorithm to train the model to perform a specific task.

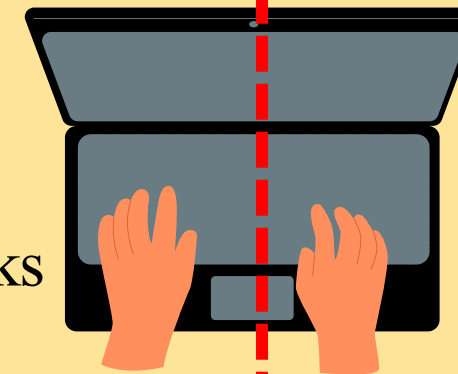The AI model is ready to receive instructions to complete tasks.

Output is generated for you to use in your operations or on screen.
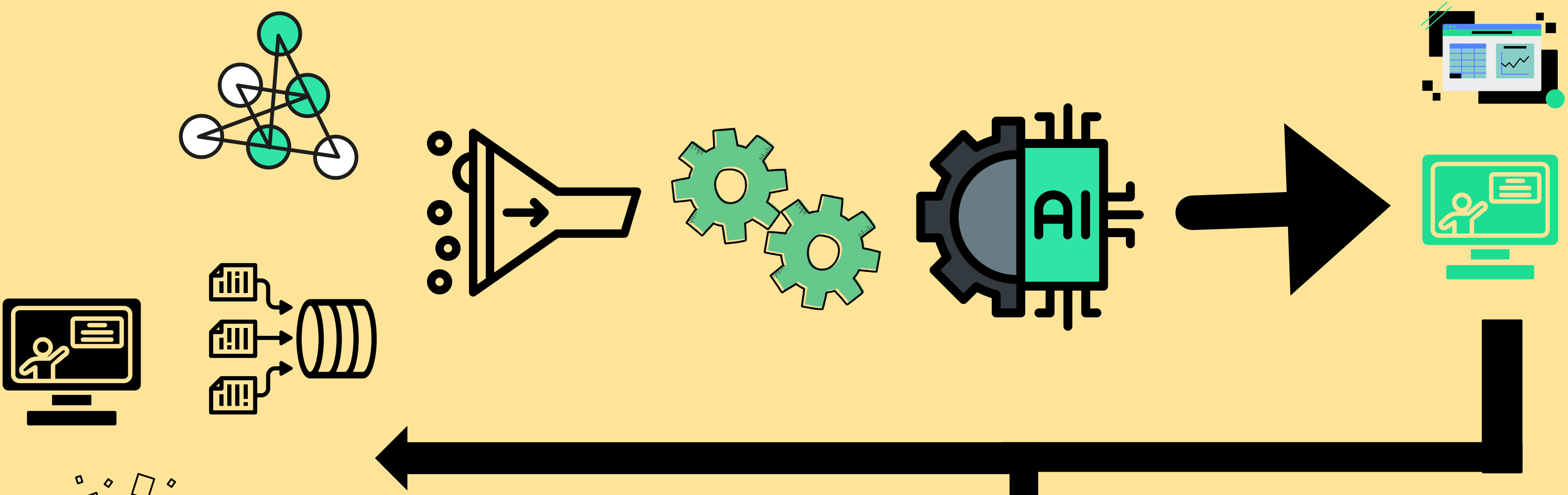
Content is prepared to be machine readable.

Content or "training data" is collected, through 'data mining' or 'data licensing'.
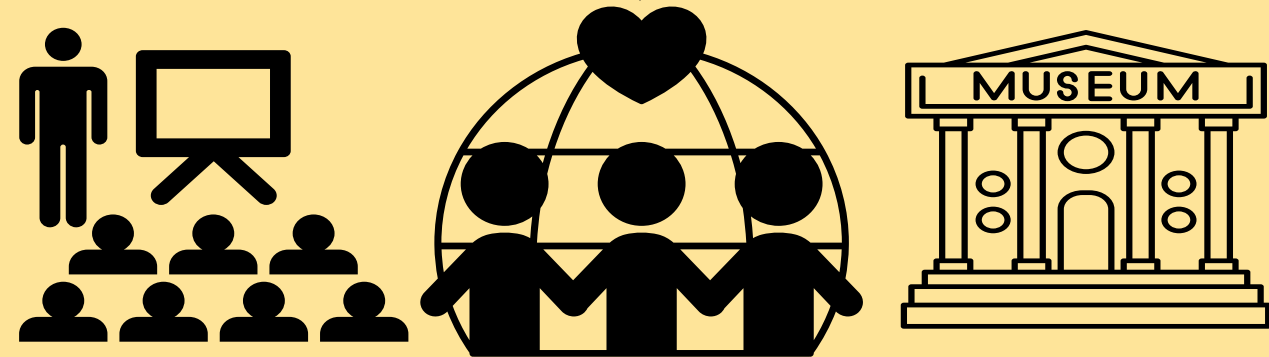
You prompt new tasks on demand

Some tasks may require that you upload their content for the AI model to ingest.
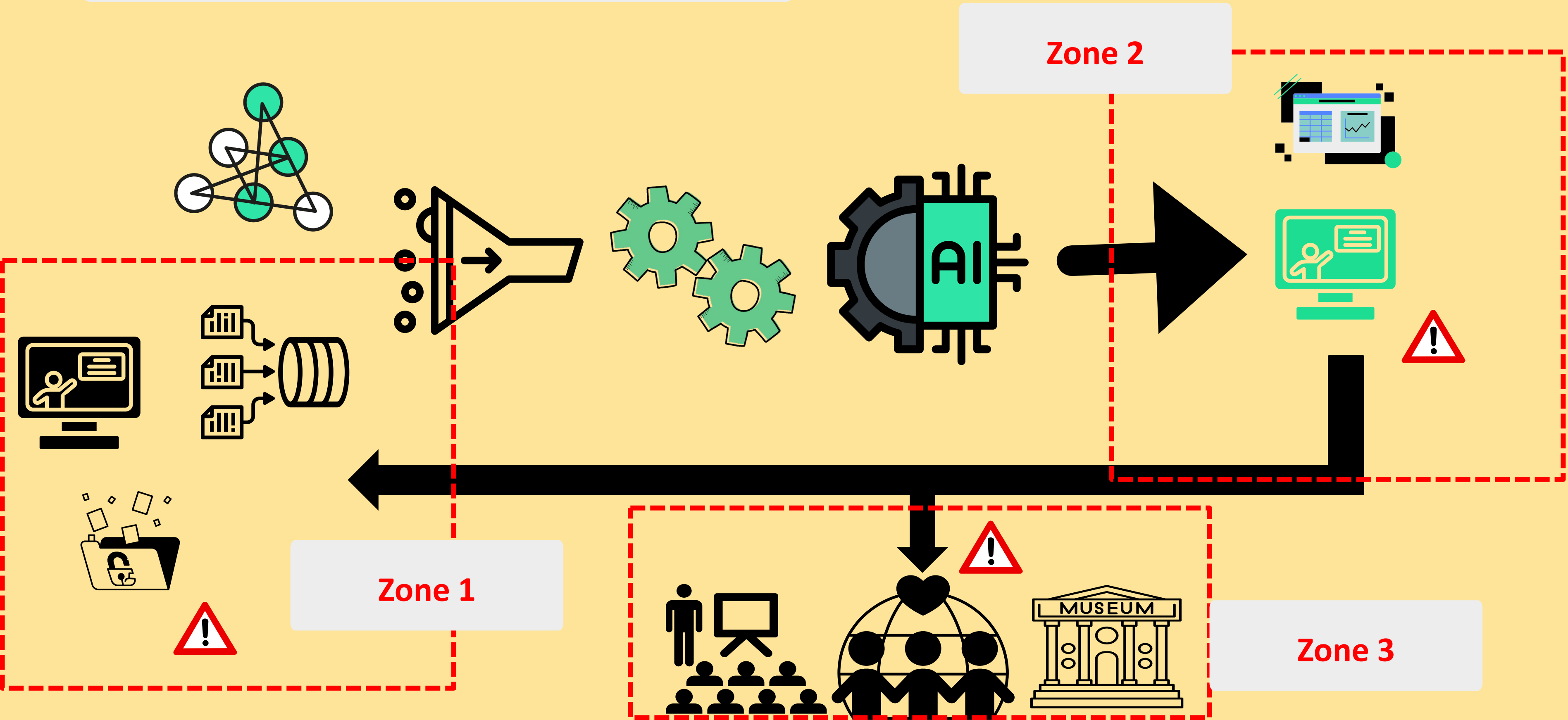
Next steps for AI-generated content

AI- generated content re- inters the data pipeline

AI- generated content is represented to or shared with audiences as ICH in and of itself or ICH derivatives, and subsequently engaged with or safeguarded by stakeholders.
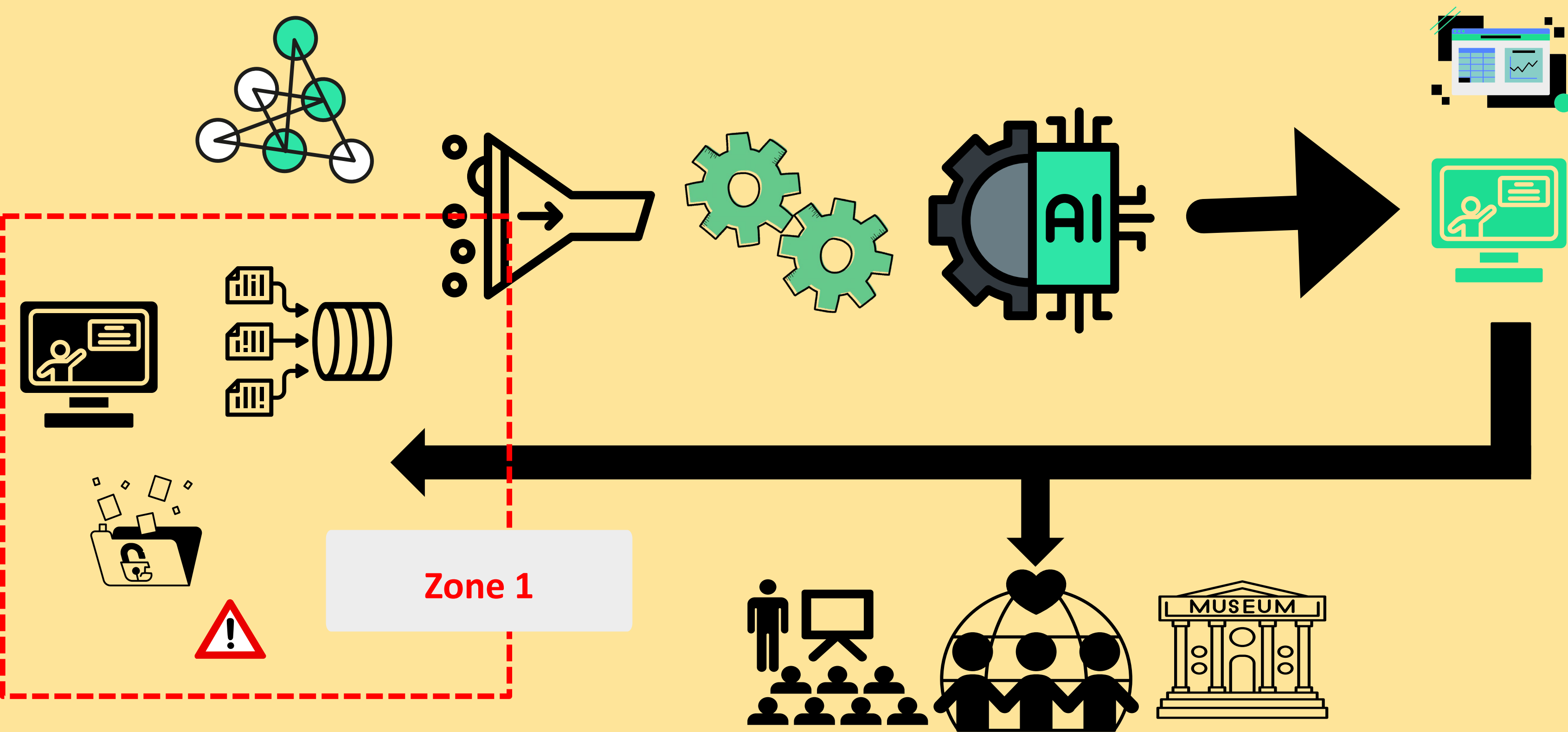
MUSEUM

Key zones of interest

Zone 2

Zone 1

Zone 3

Zone 1:
ICH ingested for AI training

Key points of risks

Zone 1

Zone 1: how ICH data is ingested by AI

What should you think about?

Determine whether it is appropriate or not for ICH content to be used for training purposes, and by whom, and under what condition? Yes / No, conditions.

1 ## Web 'crawling' or 'scraping'

A computer program known as a 'bot' mines content available online via websites and social media publications. It often makes copy of the content, and collates these copies into one large corpus of information ('dataset'). Sometimes this task is done or supervised by a human.

- Decide what and where to publish your content online
- Communicate your intentions on data mining in:
  - the metadata of your digital files, and
  - in text form next to your content, and in your terms and conditions if you have them.

2 ## Platform service conditions ('back door access')

Platform services make it a condition of their service to you that you grant permission to use ICH content uploaded onto their platform for training purposes. This is a common term of social media, server storage, file-sharing or file-editing platforms or software services.

- If you do not want your content to be mined by the companies whose services you use: check your terms of service with existing suppliers, communicate your intentions with the service provider, and request that your content be excluded from data mining purposes (also known as 'opting out'); select service providers who will allow you to do that.

3 ## Licensing deals

An AI developer forms an agreement with rightsholders and communities to access and use their content. These deals are rare but growing; few are found in the heritage sector at the moment, and ever fewer relate to ICH content.

- Get independent commercial and/or legal advice before agreeing to such an agreement. Consider partnering with other communities to increase your bargaining power.
- Grant access a limited time period, and limited purposes, instead if a one-time deal in perpetuity.
- Ask for financial and in-kind benefits in exchange for granting access to your content.

Intangible Cultural Heritage shapes AI (primarily by providing 'training' content for AI development)

Intangible Cultural heritage

Artificial Intelligence
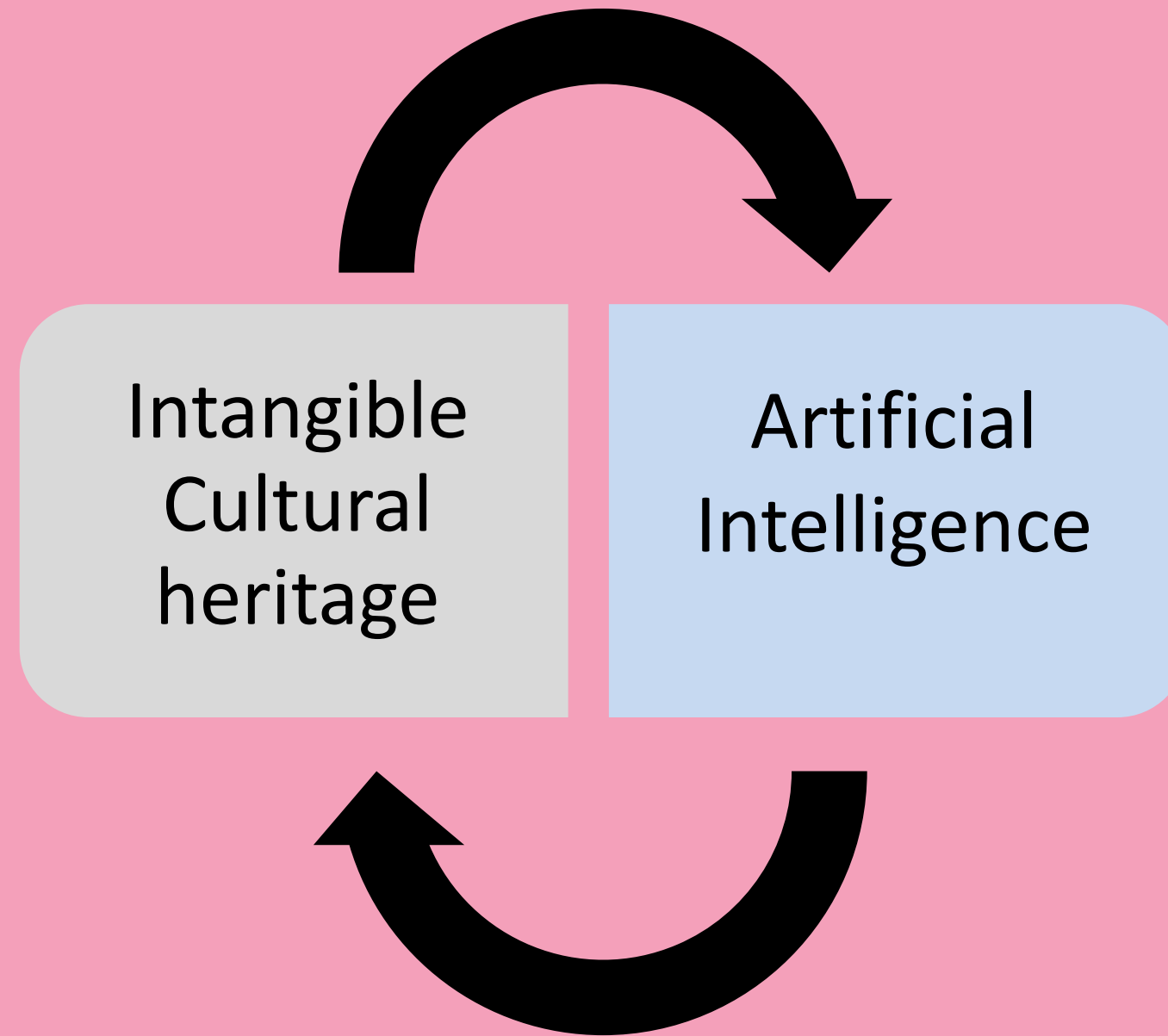
AI can impact on Intangible Cultural Heritage (primarily by providing tools to analyse, generate or share ICH)

Dr Mathilde Pavis | Pavis Consultancy

www.mathildepavis.com
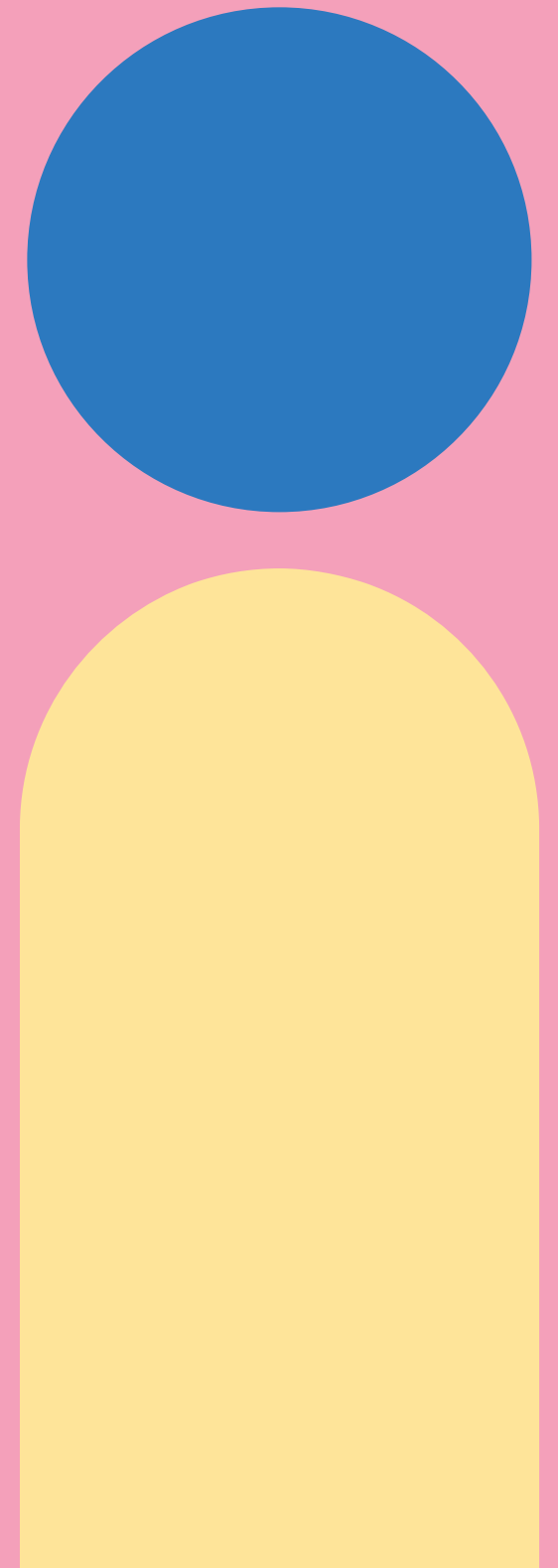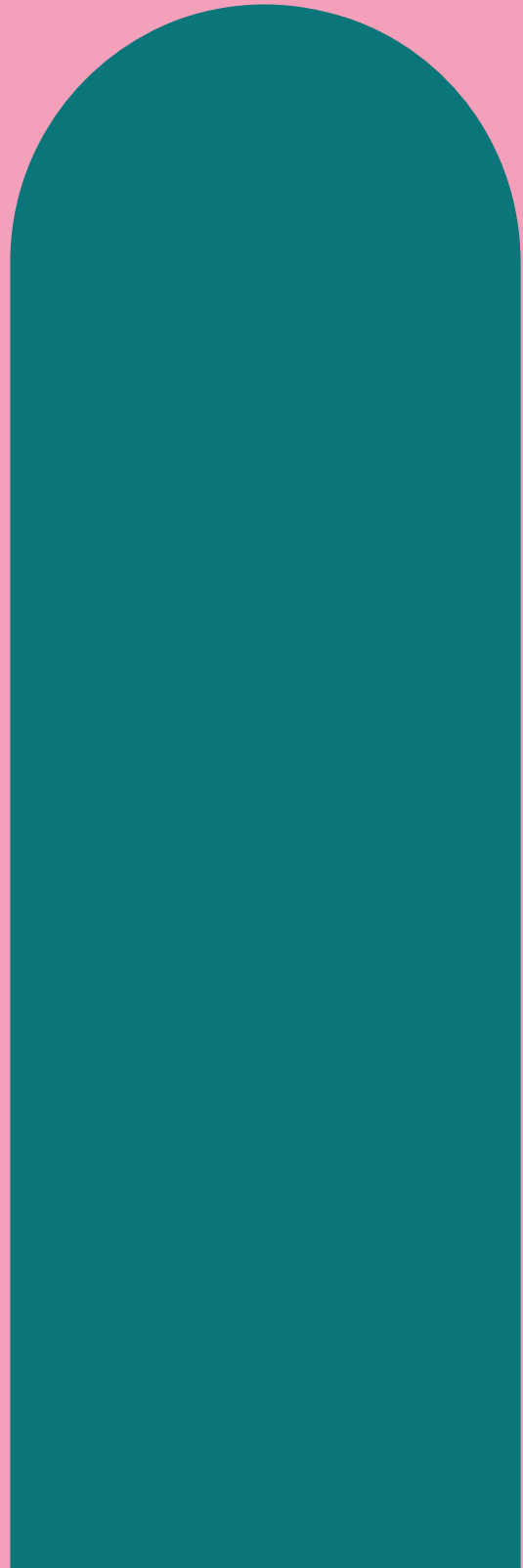
THANK YOU / MERCI

Feel free to share your questions in the chat
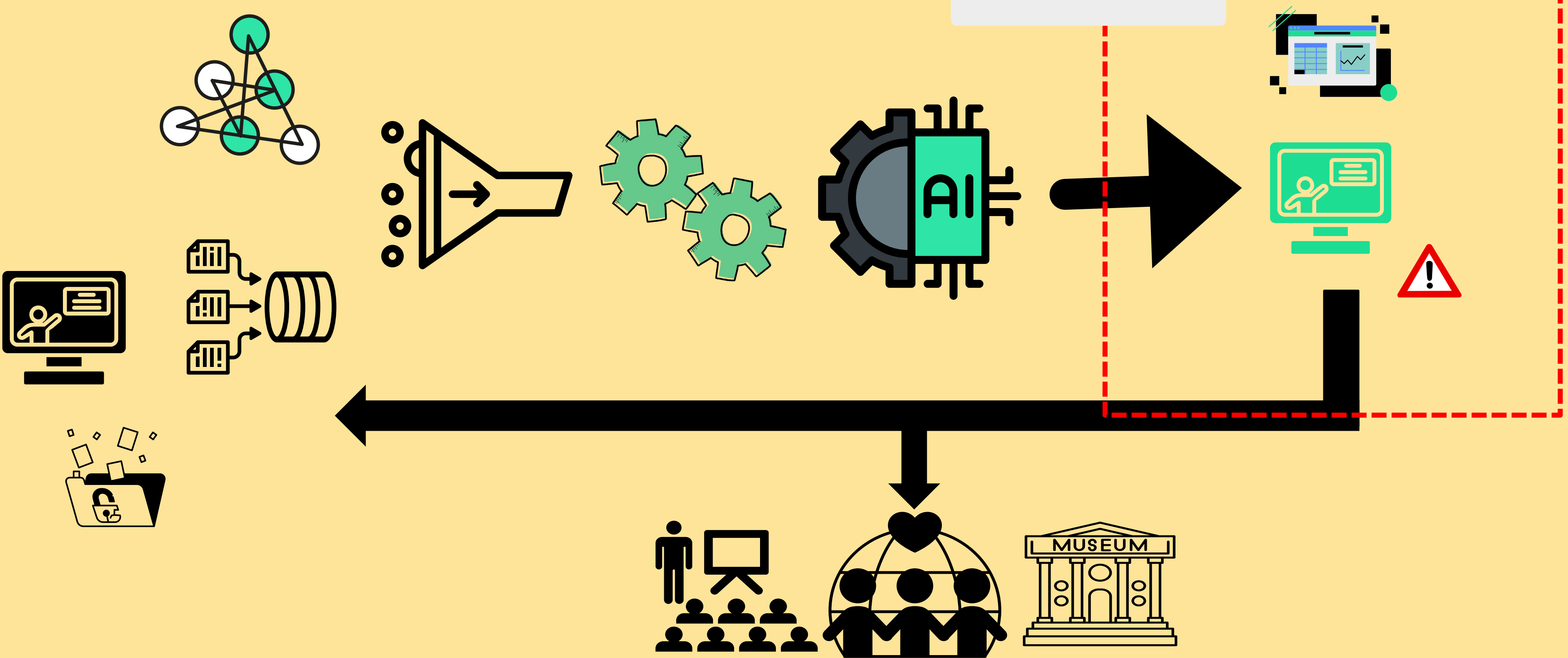
Additional slides

If you are interested

Zone 2:
Generating ICH with AI

Key points of risks

Zone 2

Zone 2 : AI-generated ICH, or ICH derivatives

Select your AI model carefully: prefer clean and specialised models

Check the T&C with your AI supplier

Select content you use on AI model carefully (ie, check the rights and get permissions)

Check outputs before using or publishing them & label your content as AI- generated

Disclose your use of AI to collaborators, clients and the audience.

Check your agreement with collaborators, community leaders, or partner allows you to use AI as you intend

Zone 3:
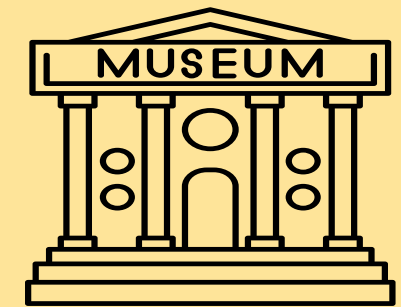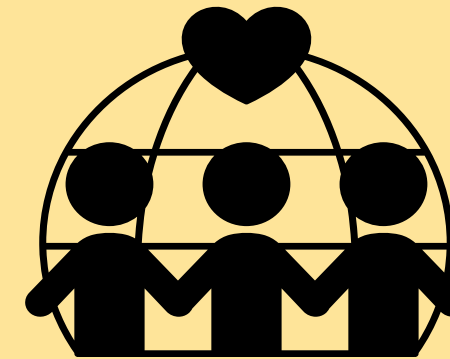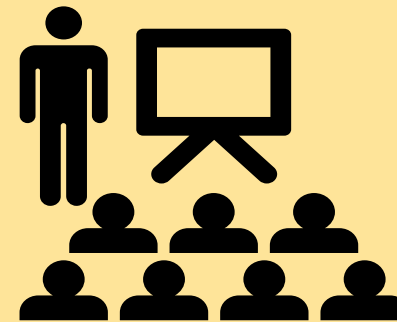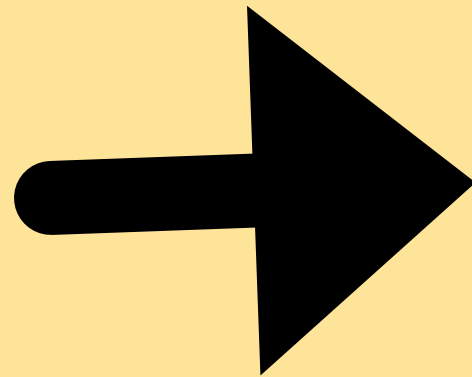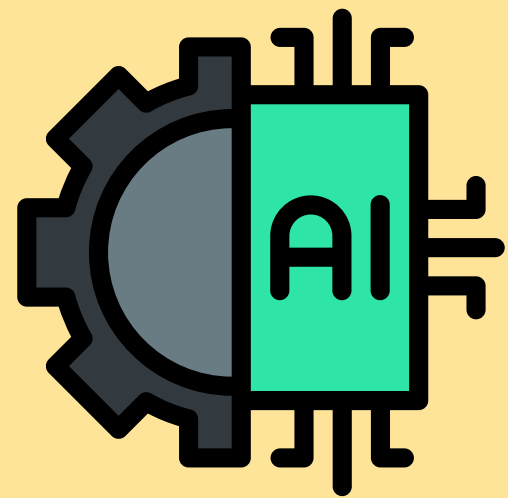Discerning between AI-generated ICH and 'traditional' ICH

Key points of risks

Zone 3
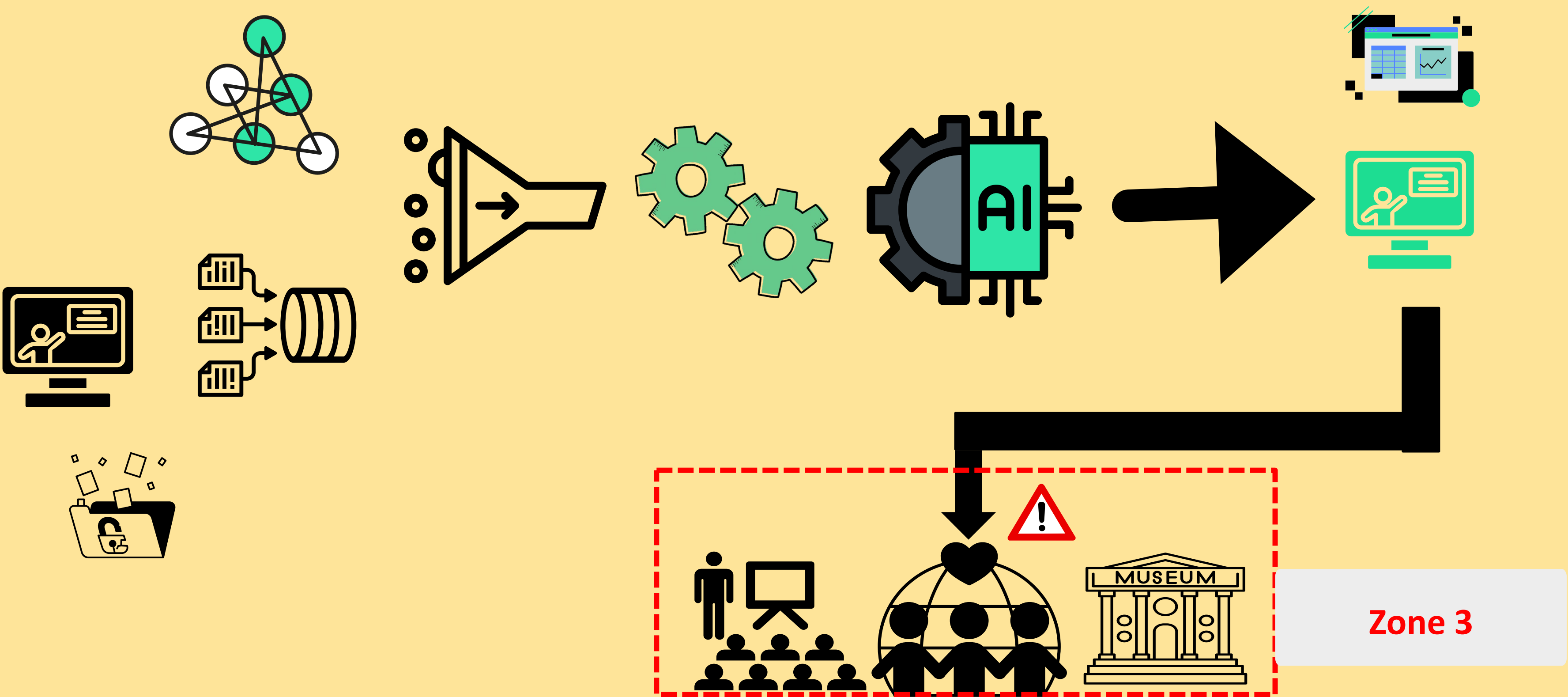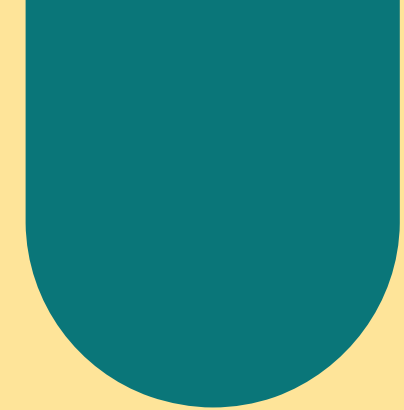
Key points of risks

Examples:
- You receive a submission to list a practice or tradition as ICH on your inventory, and the description or images have been AI-generated by the person proposing the addition.
- You want want to share knowledge about your own ICH, and want to check whether the illustration or examples you give are AI-generated.
- You are learning about your own ICH, and wonder whether the content you are engaging with was AI-informed, AI-generated, or AI-edited.

Your main risk:
- Avoiding misinformation or misrepresentation of the content you collect, list, learn from or share with others.
- Avoid recording or disseminating errors that may exist in AI-generated content and go undetected.

Zone 3: Engaging with ICH content that may be AI-generated

"Due diligence on AI"

- Ask for 'data provenance' information specific to AI

- Check the content for information in the metadata or accompanying text for disclosures on how AI was used.

- Perform your own detection checks, where detection tools exist.

- Record the results of your checks with the content, and share it with your collaborator and the audience. Label the content with this information if it is not clear as it is.