

Towards smoother research data management

—
Utilizing AI in creating research-related documents

A!



Laura Mure laura.mure@aalto.fi
Aalto University



Antti Rousi, antti.m.rous@aalto.fi
Aalto University

What I'll cover today:

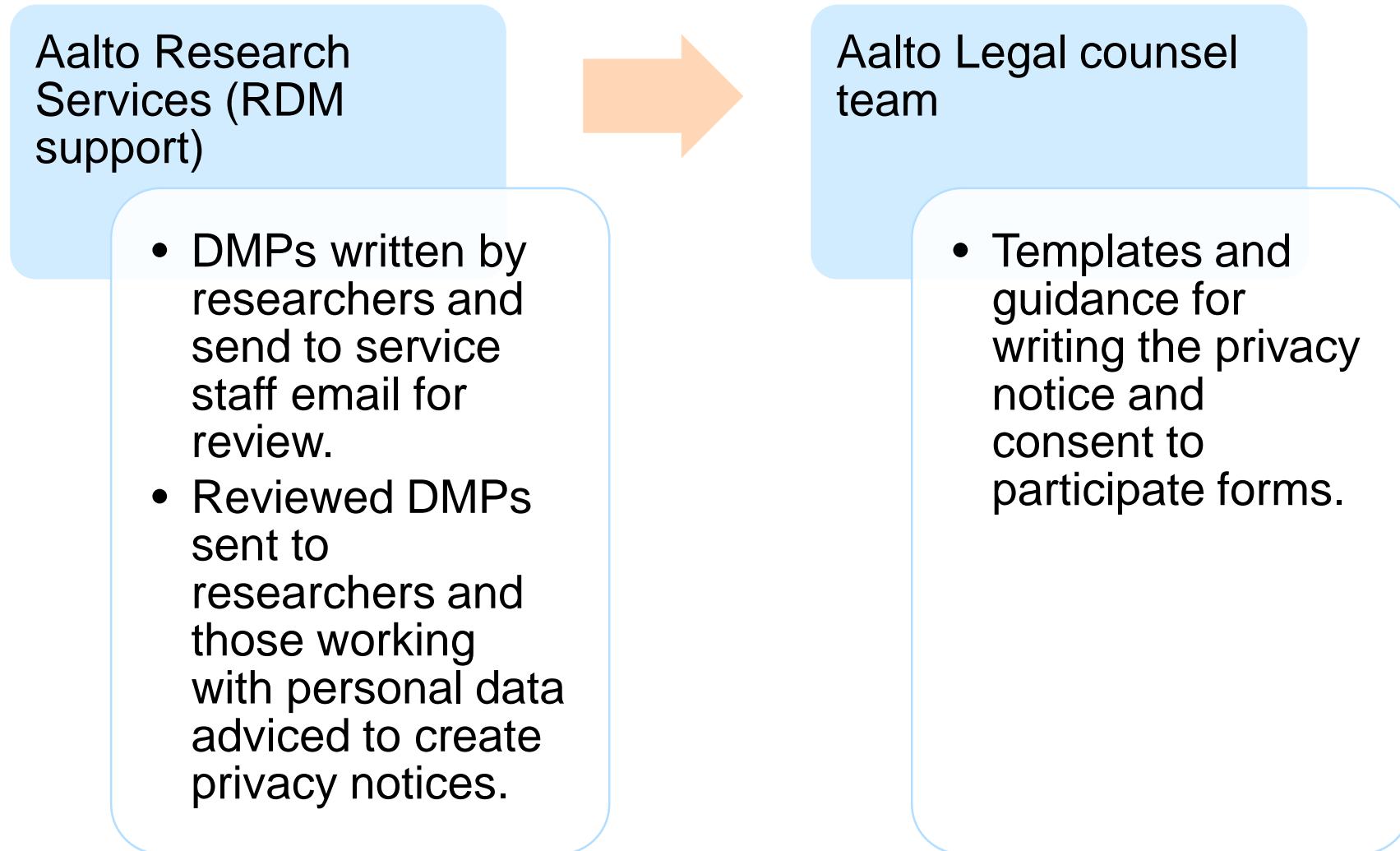


How DMPs are transformed into privacy notices using Large Language Model tool (LLM)



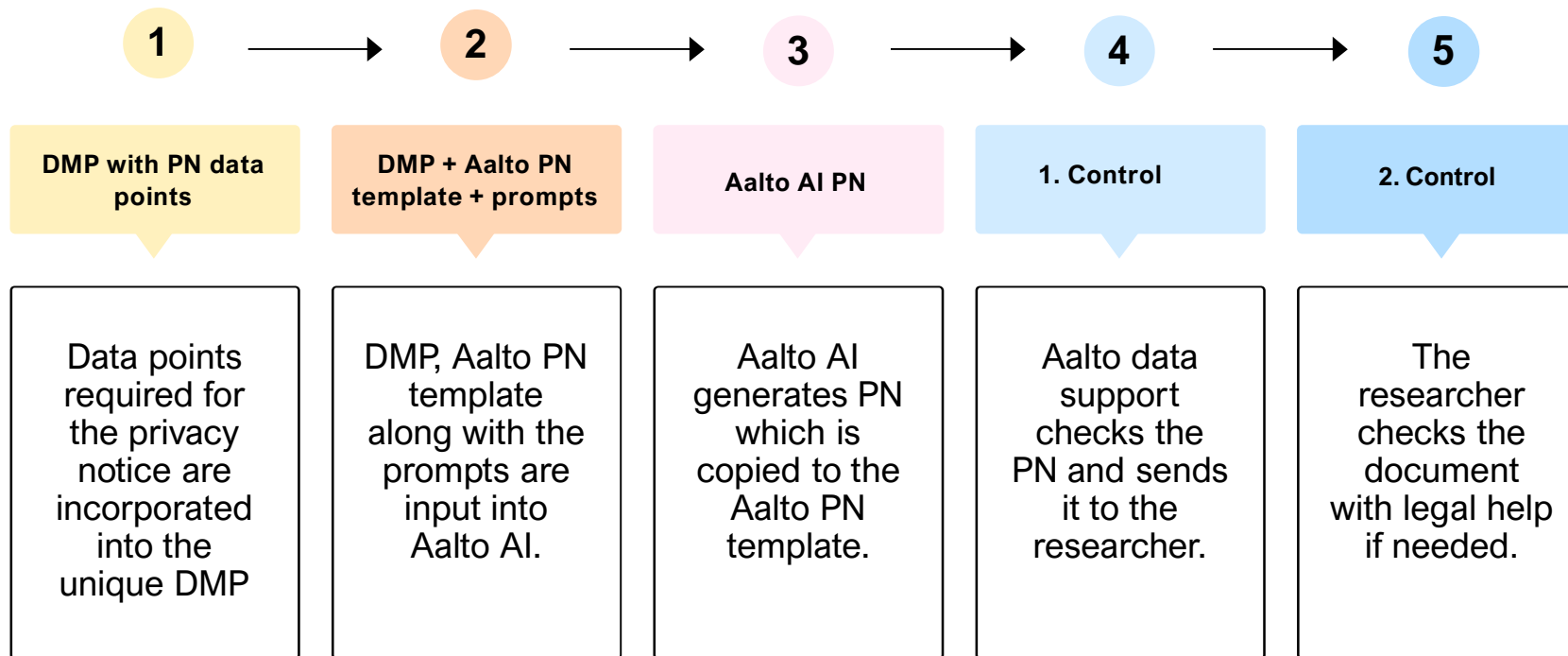
Aalto's visions for AI in research documentation

Where we began: the Aalto privacy notice (PN) process



AI PN project in a nutshell

- A pilot project to generate privacy notices using an Aalto AI tool for **Research Council of Finland (RCF)** research projects involving personal data:
 - Suitable for projects **without complex personal data issues** (e.g. no ethical review required, or special categories of data handled).
 - Goal: Reduce workload for researchers and service staff while ensuring **GDPR compliance**.
 - The first cases were processed summer 2024. Project is still on-going.



A!

LLM AI application used in the pilot: Aalto AI Assistant

- Aalto AI Assistant is a **chat-based tool** built on OpenAI's **GPT language models**, the same family that powers ChatGPT
- It is tailored for the **Aalto community** and operates within Aalto's **secure digital environment**
- The tool uses **GPT-4o (Omni)** by default -> Privacy notices were created with this model or **GPT-4 for long documents** which allow 3-5 times longer queries and also provide much longer responses.

Key benefits for this project:

- Keeps data within Aalto's digital infrastructure.
- Aalto AI Assistant can process **public, internal, and confidential data**; however, **secret content and personal data (especially special categories) are not allowed**. -> All personal data were excluded from the documents.
- **Free to use** for Research Services data support (not free though for the IT Services!)

A!

How we did it: prompts and documents

Input type	What was done	Contributor
PN	Aalto's existing template was slightly modified to include relevant data points for the AI tool.	Research and Legal Services
DMP	RCF's template was adapted to include specific sentences for AI processing.	Research Services
Prompts	Custom prompts were created to guide the AI in generating content from the documents.	Research Services with some support from Aalto Research Software Engineers

A!

Data points needed in Research Council of Finland (RCF) DMP

RCF DMP Section 1. General description of data

- Name of the template (RCF template)
- Short abstract in the DMP or research.fi
- 1.1 What kinds of data is your research based on?

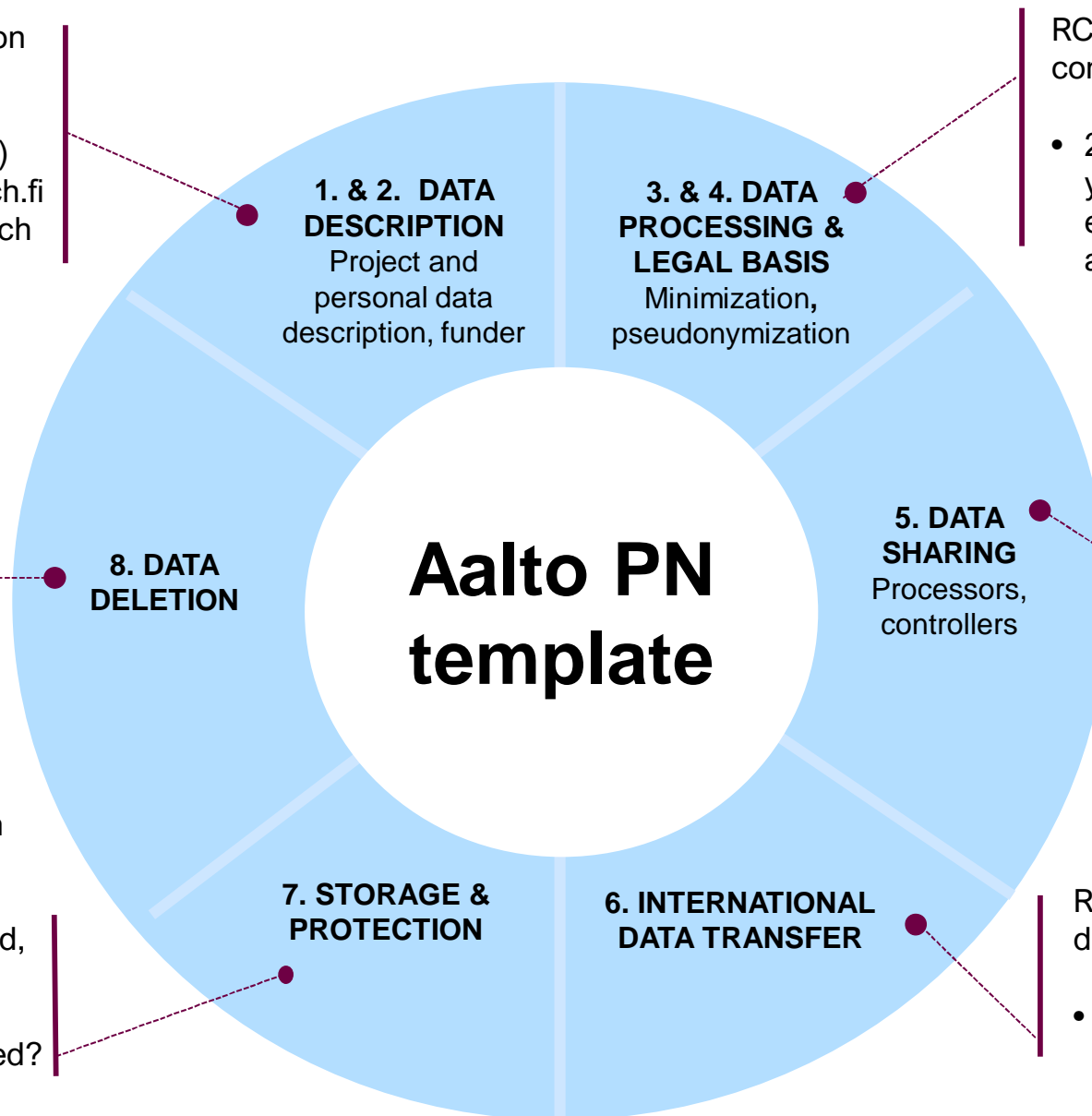
RCF DMP Section 5. Opening, publishing and archiving the data after the research project

- 5.2. Where will data with long-term value be archived, and for how long?

RCF DMP Section 4. Storage and backup during the research project and 5. Opening, publishing and archiving the data after the research project

- 4.1. Where will your data be stored, and how will they be backed up?
- 5.1. What part of the data can be made openly available or published?

A!



RCF DMP Section 2. Ethical and legal compliance

- 2.1 What ethical issues are related to your data management? (For example, GDPR and other legislation affecting data processing)

RCF DMP Section 2. Ethical and legal compliance & section 4. Storage and backup during the research project

- 2.2 How will you manage the rights of the data you use, produce and share?
- 4.2 Who will be responsible for controlling access to your data, and how will secured access be controlled?

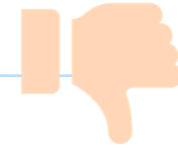
RCF DMP Section 4. Storage and backup during the research project

- 4.1 Where will your data be stored, and how will they be backed up?

Lessons learned: using LLMs to create privacy notices



- LLMs can help researchers or research support staff write a **good first draft**, so they don't always need help from legal experts for basic content.
- LLMs can assist in **translating** privacy notices into multiple languages.
- **Productivity gain:** With expert supervision, LLMs can significantly support and speed up document creation.
- **Enables meaningful human–AI collaboration:** Human expertise is essential to develop and oversee LLM-based applications.

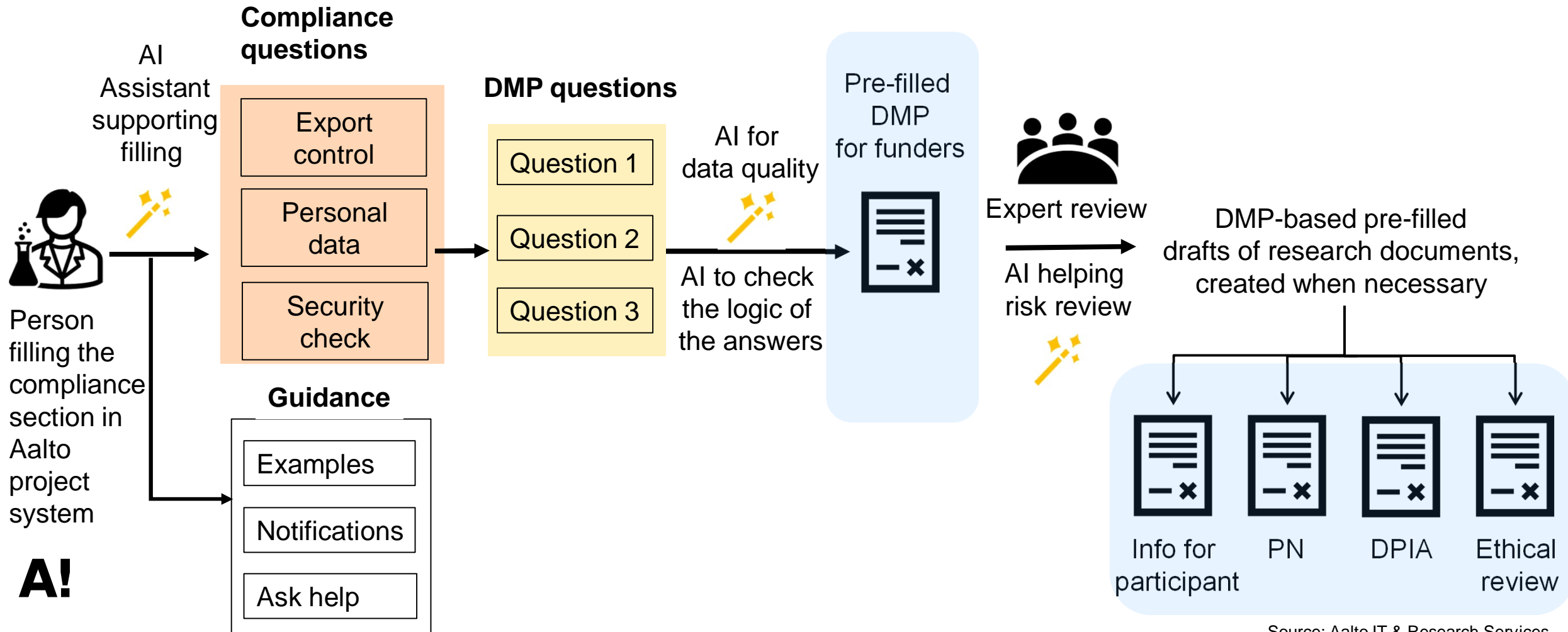


- LLMs **cannot fully automate** the transformation project data from DMPs to privacy notices.
- **Unsupervised use** is unreliable for critical documents
- Privacy notices (and other documents like DMPs) will likely be only **partially pre-filled**.
- If something isn't written in the input, LLMs won't know it — unlike humans, **they can't read between the lines**.

Future plans: project workflow planning with LLM AI help

BASIC PRINCIPLE:

All documents use the same information based on DMP and Compliance questions.
Each piece of information is entered only once.



LLM AI in research documents: A SWOT Analysis

STRENGTHS

- AI tools can draft research documents quickly, saving time.
- Once information is entered, it can be reused across multiple documents and even translated, reducing redundant work.
- The assistant lowers administrative workload for researchers and staff without requiring a major IT overhaul.

WEAKNESSES

- Outputs can be unpredictable; language models sometimes hallucinate or misinterpret inputsAI doesn't make accessible documents
- The system can only work with what you provide; it cannot infer missing details
- Needs a human to validate the final content and accessibility

SW
OT

OPPORTUNITIES

- Research institutions are experimenting with maDMPs and AI-assisted document review
- AI agents with planning and tool-calling capabilities are emerging, offering the potential to automate more routine tasks.
- Open-source models make it possible to tailor tools to local policies, languages and workflows.

THREATS

- Media hype may create unrealistic expectations; adopting AI too quickly without governance can lead to compliance issues
- Sensitive data could be mishandled if systems lack proper controls and audits
- People might be hesitant to trust AI-generated documents as legally reliable.

—

Additional material

External Funding Management for Research - *Aalto Project List (APL)*

- Aalto Project List (APL) is a platform for managing externally funded research proposals and projects in Aalto University.
- Designed for the Aalto community – researchers, service staff, and leadership
- Built on Aalto's Salesforce platform
- Supports internal approvals for research proposals and projects
- Enables secure document sharing between researchers and service personnel
- Ensures centralized documentation and version control

Links

Aalto AI Assistant

<https://www.aalto.fi/en/services/aalto-ai-assistant>

Finnish OSTRails project

<https://wiki.eduuni.fi/pages/resumedraft.action?draftId=534976212&draftShareId=30b6c1b1-ea4a-4ea8-a053-2029cc5195e6&>

RCF DMP requirements

<https://www.aka.fi/en/research-funding/apply-for-funding/how-to-apply-for-funding/az-index-of-application-guidelines2/data-management-plan/data-management-plan/>

Research Council of Finland DMP template:

https://dmptuuli.fi/template_export/254952160.pdf

A!



Kiitos
aalto.fi