

Incorporating AI into Statistical Standards: Enhancing GSBPM with (Generative) AI

UNECE | ModernStats World Workshop, October 2024

Riitta Piela, Statistics Finland

Concepts

- AI, includes both “traditional” and generative AI
- Responsible AI, refers to the development, deployment and use of artificial intelligence in a way that ensures fairness, transparency, inclusivity, security, privacy, ethics and accountability.
- MLOps, (Machine Learning Operations) is a set of practices, tech components and roles, that combine machine learning (ML) with DevOps principles to streamline the development, deployment and management of ML models in production. It ensures reliable, transparent and repeatable model performance.
- LLMOps, (Large Language Model Operations) is an extension of MLOps, specifically focused on managing and operationalizing large language models like GPT.



AI's role in GSBPM

- **Three AI Roles**

- Autonomous AI: full automation with minimal human intervention.
- Collaborative AI: partial automation with human oversight.
- Supportive AI: AI as an assistive tool enhancing human efficiency.

- **Criticality of AI's role in business**

- AI applications can also be classified based on how critical the AI-executed tasks are to the core functions of official statistical production.
- In highly critical tasks, AI's operation must be fully transparent, and the quality of predictions must be continuously monitored (adhering to MLOps principles). This is especially challenging with generative AI, where ensuring prediction quality and transparency can be more difficult (LLMOps).
- This approach is conceptually linked to the EU's AI Act, though it does not directly align with the risk-based classification in the Act.

- **Challenges:** Transparency, reproducibility and ensuring compliance with Responsible AI principles remain critical.



Integrating AI into GSBPM

- Opportunities
 - Possible **automation** of tasks within GSBPM.
 - Assisting in process and data quality improvement across different phases within GSBPM.
 - AI can save employees' time (supportive AI), enabling them to focus on developing innovative services and exploring new opportunities.
 - Faster, more efficient statistical production.
 - Activating global collaboration with producers of official statistics...
- Challenges
 - Transparency and reproducibility concerns, especially with generative AI, due to the complex nature of these models and their reliance on vast, sometimes opaque, datasets (LLMOps is the key, but not a solution).
 - Ensuring Responsible AI principles.



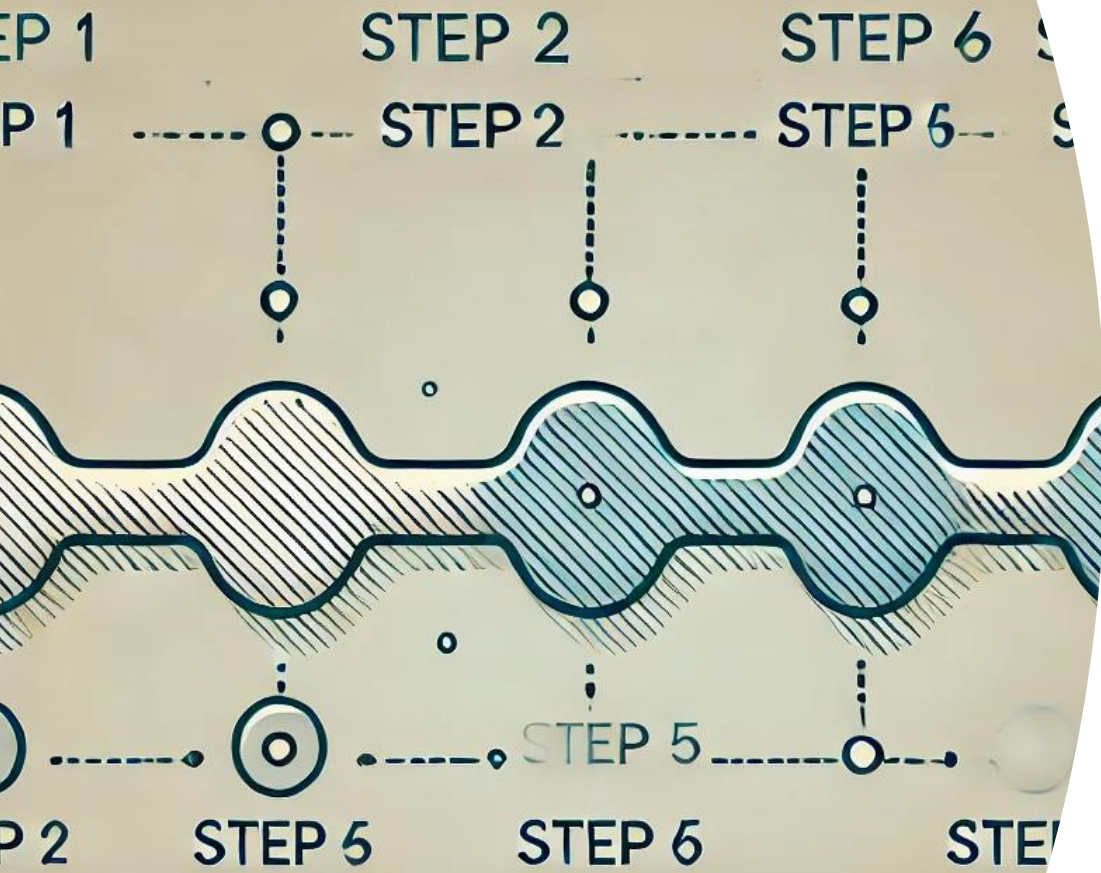
Preliminary Working Model: steps for AI implementation (1/2)

- Step 1: Problem identification
 - Identify a problem or task within GSBPM.
 - Determine if AI, including supportive AI, can provide a solution (e.g., how supportive AI has been used for a similar task).
 - Evaluate the ROI of AI possible usage.
 - Check for existing solutions (or whether a statistical office has already identified and is researching the issue) within the AI-GSBPM platform.
- Step 2: Feasibility assessment
 - Assess if AI, including supportive AI, is a viable solution.
 - Define the role of AI: autonomous, collaborative or supportive (and how critical...)

If such a service/idea does not exist →

- Step 3: Data discovery and preparation
 - Identify and gather relevant data (for training, fine-tuning, RAG purposes)
 - Ensure data quality and relevance
- Step 4: Model development or adaptation or...
 - Develop or adapt a machine learning model or implement a supportive AI solution observed from other use cases.





Preliminary Working Model: Steps for AI Implementation (2/2)

Step 5: Responsible AI review

- Ensure compliance with Responsible AI principles (fairness, transparency, reproducibility, security).

Step 6: Stakeholder analysis

- Analyze the AI model's (or supportive AI's) solution's impact on statistics and stakeholders.
- Gather feedback and use it for improvement.

Step 7: Integration with GSBPM

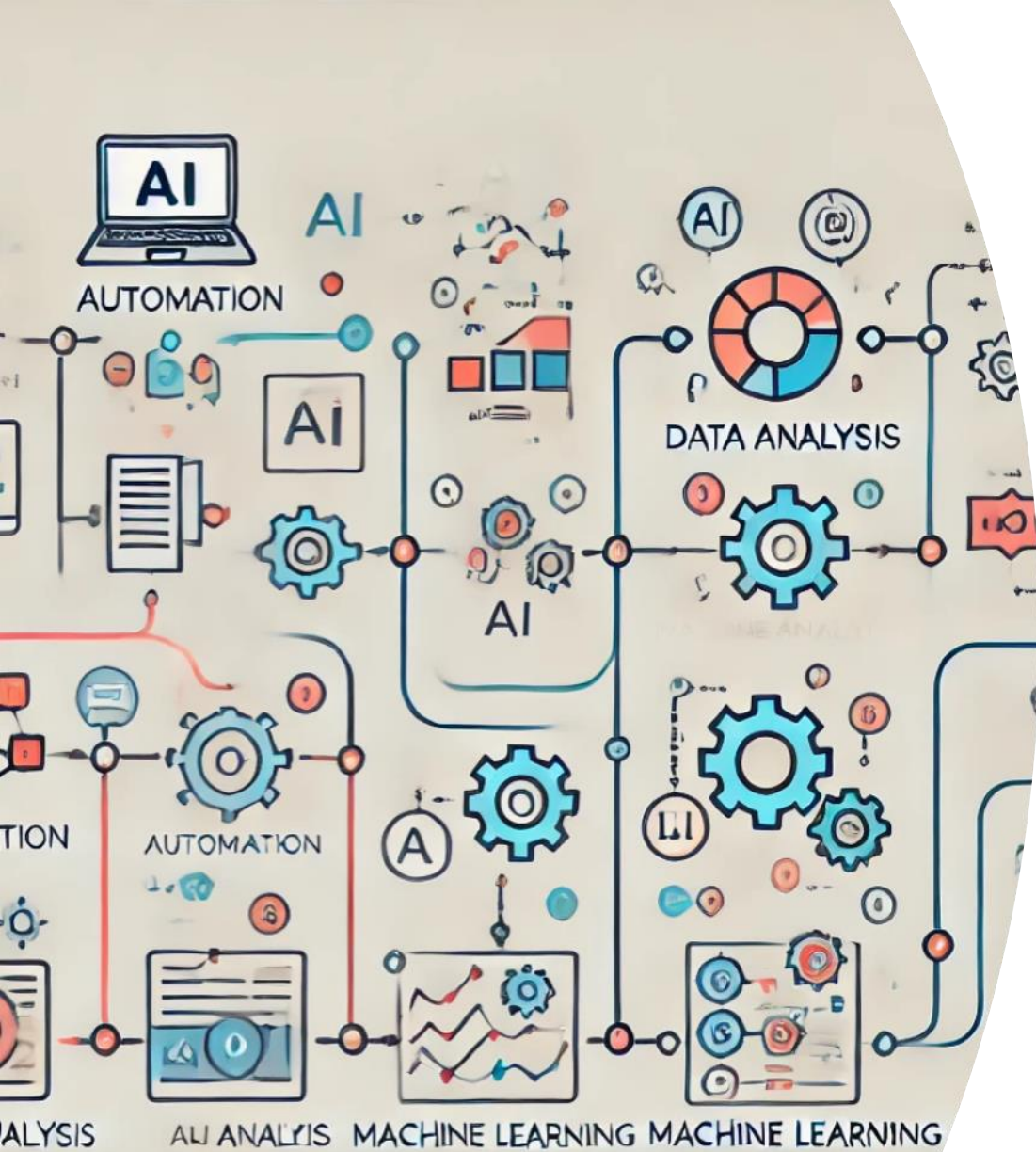
- Map the AI solution (or supportive AI implementation) to the relevant GSBPM phase.
- Share the solution on collaborative platforms for wider accessibility (for example stored in a form of **Model Card**).



Collaborative platforms for AI solutions

- Platforms
 - Encouraging open communication and sharing of AI solutions across statistical organizations.
- Examples
 - "Awesome Official Statistics Software" repository: [Sharing tools and methodologies for statistical production.](#)
 - Generative AI for Official Statistics Repository: [Generative AI for official statistics repository \(unece.github.io\)](#) Focused on generative AI use cases in official statistics.
- Future considerations
 - Develop a platform specifically for AI + GSBPM solutions to enable sharing, **collaboration** and discovery of AI use cases.
- Key question
 - What platform features would enable the description and sharing of functions based on supportive AI roles, where AI assists without necessarily involving machine learning models?





AI and GenAI in GSBPM Phases (examples)

- **Specify Needs**
 - Example: GenAI analyzes stakeholder feedback and identifies trends in needs. AI chatbots gather continuous feedback, improving real-time engagement (supportive AI).
- **Design**
 - Example: AI helps optimize survey designs and data collection instruments (collaborative AI).
- **Build**
 - Example: AI adapts survey instruments in real-time based on respondent behavior and simulates the statistical process to detect inefficiencies (autonomous AI).
- **Collect**
 - Example: AI assists in automating the sampling process, ensures optimal sample size and adapts survey questions in real-time to improve relevance and data quality (supportive AI).



AI and GenAI in GSBPM Phases (2/2)

- **Process**

- Example: AI supports data integration, anomaly detection and intelligent imputation of missing data. Generative AI suggests corrections by detecting inconsistencies in datasets (collaborative AI).

- **Analyze**

- Example: AI detects patterns, creates intuitive visualizations and assists in anomaly detection to validate outputs, enhancing accuracy and depth of analysis (collaborative/supportive AI).

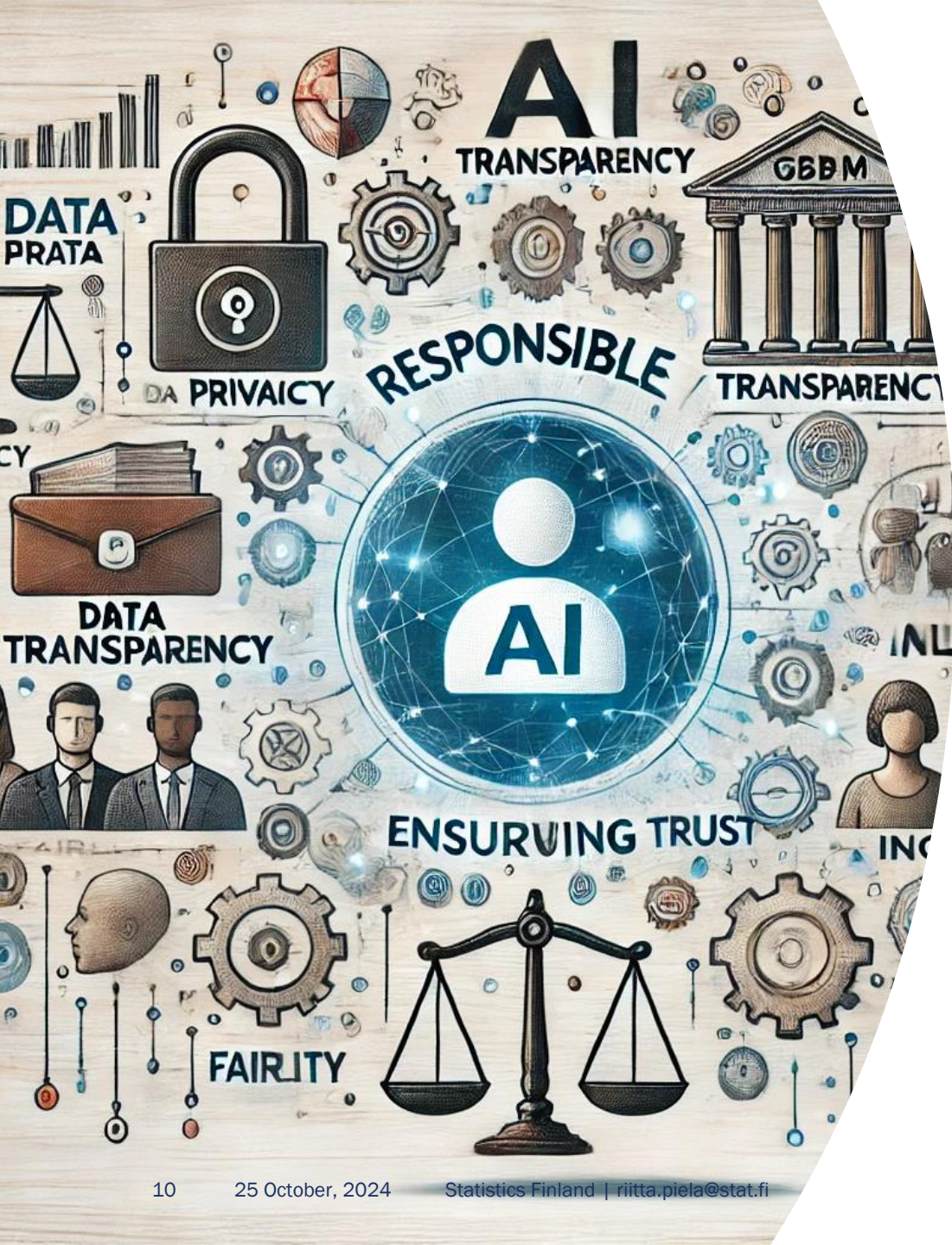
- **Disseminate**

- Example: AI generates personalized reports, automates data visualization for different stakeholders and AI chatbots provide instant query support on disseminated statistics (supportive AI).

- **Evaluate**

- Example: AI assesses the quality of statistical outputs using predefined metrics, tracks quality issues throughout the process and identifies root causes using production chain metadata (collaborative AI).





Responsible AI in GSBPM: ensuring trust

- Key Issues: data privacy, transparency, bias and fairness are critical when integrating AI into GSBPM phases.
- Responsible AI Framework (ADSaMM groups's framework): promotes transparency, reproducibility, inclusivity... → are essential for maintaining public trust in AI-enhanced statistical production process.
- Current efforts: ADSaMM group's initiative on advancing Responsible AI
 - Publication: Responsible AI Framework for official statistics.
 - Program: Advancing Responsible AI through a targeted onboarding program for diverse stakeholders in statistical production.





Future of statistical production with AI

- Vision: AI-driven statistical processes
 - Autonomous AI, Collaborative AI and Supportive AI will collectively revolutionize statistical production by automating processes, improving decision-making and assisting staff with complex tasks.
- Potential: AI-driven automation and real-time data processing will significantly speed up the production of statistics, enabling faster response times to emerging societal needs.
- Automation's Role: redefining the production process with automated, AI-driven workflows (AI agents replace the human-in-the-loop).

Conclusion

- Traditional AI and Generative AI are transforming statistical production across all phases of GSBPM.
- Opportunities: AI enhances efficiency, automates routine tasks and allows staff to focus on innovation, creating new services and improving statistical production.
- Challenges
 - Transparency, reproducibility and ensuring compliance with Responsible AI principles remain critical, particularly for generative AI.
 - Lack of effective platforms for sharing solutions leads to wasted time solving problems that may have already been addressed elsewhere.
- Next Steps: Continue developing **AI guidelines, platforms for sharing solutions** and embedding Responsible AI principles to sustain trust and integrity in statistical production.
- We want to combine our expertise and work together to create solutions that help us advance statistical production and data generation to a new era.



A woman with shoulder-length, wavy red hair is smiling warmly at the camera. She is wearing a light blue button-down shirt. In the background, a large presentation screen displays the text "Thank you!". The setting appears to be a conference or meeting room with an audience seated in the foreground and two men standing at a podium in the background.

Thank you!

(All images generated by DALL-E)