Summary of "*AI Agency Risks and their Mitigation through Business Process Management: a Conceptual Framework.*" - Anna Sidorova, Dana Rafiee

Section 1: Introduction

Business Process Management is a diverse research field. It emerged at the intersection of three process management traditions: quality control/scientific management, strategic management, and information technology.

Objectives:

- What factors increase the risk of the agency risks in organizational use of AI?
- How can BPM help address the AI agency risks?

Section 2: Theory & Background

In 1955, "artificial intelligence" has been coined and it had significant growth ever since, leading to self-driving cars, enhanced fraud detection, and AI beating humans in various games. Although current AI is task-specific, progress in representation, transfer, and reinforcement learning is advancing towards **Artificial General Intelligence (AGI)**.

Agency theory examines cooperation between different goals or risk attitudes, focusing on the principal-agent relationship. Governance methods focus on agency problems, aligning activities with stakeholder objectives. Metrics measurement, compensation contracts, and formalization/automation improves goal alignment and transparency, reducing information asymmetry.

Business Process Management (BPM) sees organizations as collections of processes. These processes are defined, analyzed, implemented, and continuously improved. Each process has a number of activities with an outcome, often structured with defined flows, inputs, and outputs via workflow automation but, they lack uniqueness. Unstructured processes like product development and strategic planning involve varied activities and provide a competitive edge due to their complexity. Business Process Management (BPM) classifies processes into transactional, development, enabling, and governing. Despite inputs, outputs, flow, and actors, BPM overlooks the resource usage. In these processes, roles are based on specialization and that can be predefined or flexible, with the possible agency risks that can be mitigated through compensation plans.

Section 3: Conceptual framework

AI agents perceive, act, and optimize their actions based on goal optimization, influenced by environmental understanding, either pre-existing or learned.

Organizations apply AI to enhance stakeholder value, often in specific business units for different objectives. AI goals should coincide with business objectives, but biased learning can cause goal volatility.

Tech firms and startups develop AI tools for their skills and data, aiming for profit and value because market competition and client preferences affect customization.

Sidorova, A. (2019, January 8). AI Agency Risks and Their Mitigation Through Business Process Management: A Conceptual Framework. https://scholarspace.manoa.hawaii.edu/items/20969086-e218-4ea6-967a-675df718d341

Consumer AI products like Amazon Echo gather user data for training. As data value decreases, providers might promote third-party products, reducing goal alignment for off-the-shelf AI. Alignment depends on the artifact's origin, contracts, development specificity, and data sources.

Transparency in AI operations is a pressing cncern due to several factors:

- The complexity of advanced machine learning (ML) algorithms like deep neural networks, makes them difficult to interpret. Traditional statistical models were more interpretable. Explainable AI (XAI) initiatives aim to address this issue.
- AI's ability to gather data directly from the environment, facilitated by the Internet of Things (IoT) and advancements in representation learning, further complicates transparency. As AI brokers agreements with devices and processes unstructured data, human oversight over data access reduces.

Overall, transparency is influenced by learning algorithms, the complexity of AI artifacts, data gathering abilities, and the incorporation of explanation systems within the artifacts.

AI risks increase with reduced transparency and misaligned goals. Risks are minimal with simple, in-house ML solutions for specific tasks. However, as AI sophistication and data access increase, so does the risk, with potential issues like ill-defined utility functions or biased data learning. Outsourcing AI development increases these risks due to loss of control over utility functions and training data. Mitigation includes decision model analysis and output control.

A few **use cases** of AI can be used in money-laundering schemes, ATM cameras can detect and analyse faces, segmentation of clients to personalise offers.

AI artifacts, acts as actors in structured or unstructured processes, thus increasing the risk of agency problems with reduced human supervision. To manage risks, organizations need to improve transparency and implement controls. However, the identification of AI in processes is difficult due to lack of specific notations and AI's integration into third-party software. Hence, it's recommended to model AI components in process maps for better risk assessment and stricter AI governance.

AI artifacts evolve and require specific development within organizations. Unlike traditional IT assets, they are data-based, needing accurate and verified data. Separate processes may be needed, including auditing data sources and continuous retraining. As independent learning grows, their development processes resemble human resource management. Organizations with specialized AI processes are suggested to be more successful in assessing and mitigating AI risks than those using traditional IT processes.

Section 4 & 5: Discussion / Conclusion

This paper highlights the need for AI governance research to tackle risks in BPM with studies policies and regulatory measures to enhance transparency, and developing norms for AI development. The proposed framework aims to mitigate AI risks, underlining the role of information systems in shaping the future.