AI Governance in the System Development Life Cycle: Insights on Responsible Machine Learning Engineering

Carolina Dias
Claudio Fortier

O que veremos hoje

- Introdução
- Background
- Materiais e Métodos
- Descobertas
- Discussões
- Conclusões

Introdução

- O **crescimento acelerado** do uso de soluções de ML com modelos caixa preta acendeu o alerta para a necessidade de governança de IA.
- A **governança de IA** fornece normas, políticas e instituições globais para melhor garantir o desenvolvimento benéfico e o uso de IA avançada.
- Governança de IA institucional e de desenvolvimento/ciclo de vida de sistemas.
- Novos papéis, como cientistas de dados e engenheiros de ML, precisam ser **integrados** ao ciclo de vida das aplicações.
- Garantir **explicabilidade** e **trilhas de auditoria** exigem a implementação de governança de IA.

Pergunta Principal da Pesquisa

"Quais são os principais problemas e decisões que os projetos de desenvolvimento de ML enfrentam durante seu ciclo de vida em relação à governança de IA?"

Background (SDLCs* e Desenvolv. de IA)

- O **DevOps** tornou-se o padrão para desenvolvimento de software.
- A **engenharia de software** em si precisa de adaptação quando se trata de IA e ML.
- O trabalho dos cientistas de dados é algo mais experimental e difícil de encaixar no SDLC.
- **Principais obstáculos para os cientistas de dados:** tomada de decisão com os clientes; teste e garantia de qualidade; e depuração.
- **Fatores relevantes:** modelos de ML avançados são caixas pretas; modelos de ML são probabilísticos; o desenvolvimento de ML é orientado por dados; modelos de ML não podem ser ajustados depois de treinados; e comparar qual modelo é o melhor não é uma tarefa trivial.

*SDLC: System Development Life Cycle - Ciclo de Vida do Desenvolvimento de Sistemas

Background (Governança de IA)

- **Governança organizacional:** garantir que o uso de tecnologias de IA por uma organização esteja alinhado com as estratégias, objetivos e valores da organização; cumpra os requisitos legais; e atenda aos princípios de IA ética seguidos pela organização.
- **Modelo em camadas:** técnica, ética, social e legal.
- A governança das fontes de dados em todas as etapas é essencial.
- Aumentar a **transparência** e **explicabilidade** é um dos principais focos da governança de modelos
- A governança de IA deve atuar também a nível de sistemas onde as soluções de ML são acopladas, protegendo-os de erros e comportamentos indevidos.
- Muitas vezes é preciso que a governança se dê por interferência humana.

Materiais e Métodos do Estudo

- **Design de entrevistas com especialistas:**
  - Recrutamento de especialistas na área de IA e desenvolvimento de software para entrevistas em profundidade.
  - Entrevistas semi-estruturadas de 30 a 60 min.
  - Foco no suporte conceitual do desenvolvimento de sistemas de ML na forma de modelos SDLC e como a governança de IA se relaciona com esses processos.

Materiais e Métodos do Estudo

- **Coleta de dados e entrevistados:**
  - Entrevistas remotas no Zoom.
  - 17 entrevistados, sendo 12 da indústria e 5 da academia.
- **Análise de dados:**
  - Familiarização com os dados.
  - Codificação dos dados.
  - Formação de uma estrutura de dados que descreva as entrevistas.

Materiais e Métodos do Estudo

- **Perfil dos entrevistados:**

ID	Current employment	Organization	Experience
1	Data and security specialist	Large public sector company	5+ years
2	Professor (AI-focused)	Large public University	15+ years
3	Senior data specialist	Medium private software consulting company	20+ years
4	Senior insurance mathematician	Large private insurance company	5+ years
5	Research Fellow (AI-focused)	Large public University	15+ years
6	Professor (SDLC-focused)	Large public University	20+ years
7	System architect	Large private corporation	5+ years
8	Senior software developer	Large private software consulting company	10+ years
9	Professor (SDLC- and AI-focused)	Large public University	20+ years
10	Competence lead	Medium private software consulting company	10+ years
11	Professor (AI-focused)	Large public University	20+ years
12	Director of software and services	Medium private corporation	20+ years
13	Chief technology officer	Medium private corporation	10+ years
14	Software expert	Large private corporation	20+ years
15	AI consulting expert	Large private software consulting company	10+ years
16	Data group lead	Large private software consulting company	10+ years
17	Partner in an AI company	Small private AI-focused company	20+ years

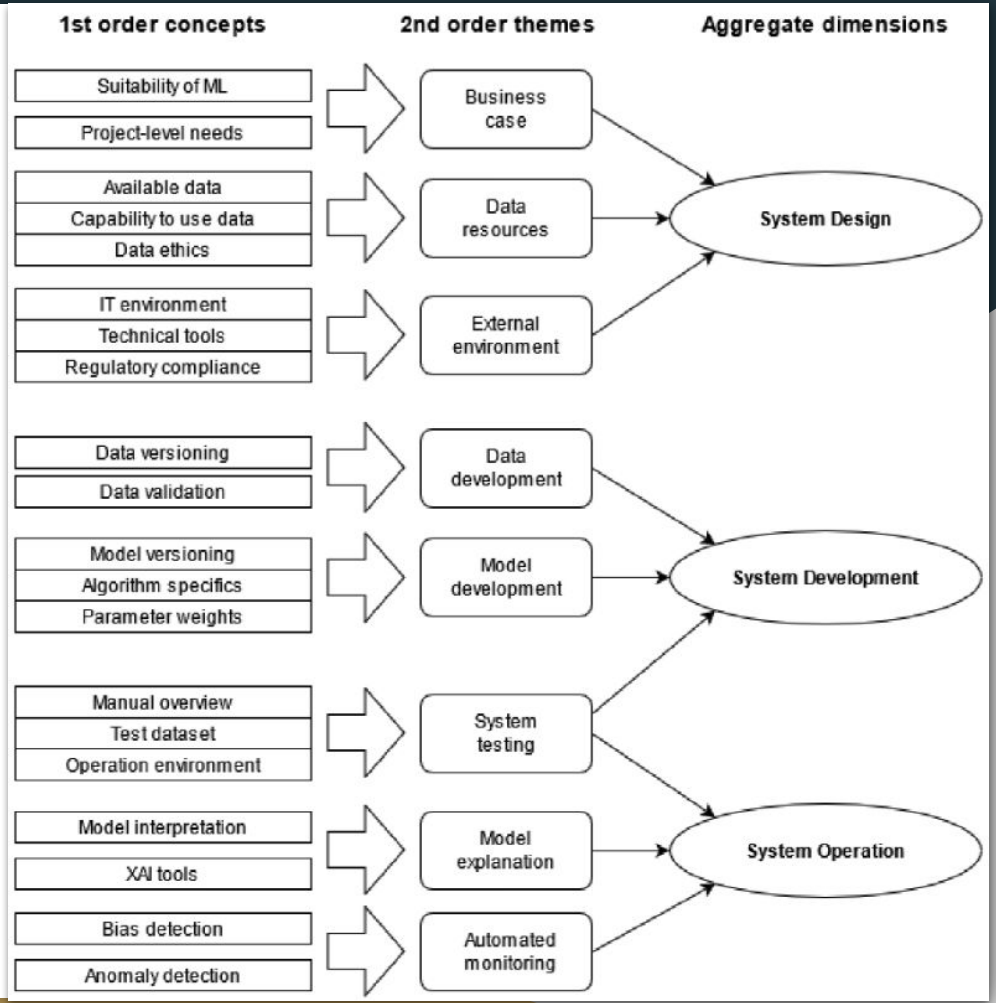
Materiais e Métodos do Estudo

- **Dados das entrevistas:**

Code	Example from the interviews
Suitability of ML	*"In fact I think our customers are oftentimes more focused on the technical implementation whereas we are still thinking about whether ML is even a suitable tool for their business case"*
Project level needs	*"first there is a need to think about the [project's] data and analytics strategy at a very high level"*
Data sources	*"ML models are closely tied to the data they are trained with."*
Capability to use data	*"(–) whether we can even use the data we have in the first place."*
Use environment	*"Training a model with bogus data gives no indication of whether it works in the real world environment."*
Technical tools	*"we can use tools such as Azure Machine Learning or Amazon Sagemaker."*
Regulatory compliance	*"Where data is stored matters from the perspective of, say the GDPR legislation."*
Data versioning	*"for reproducability in ML, data versions used for training are tracked."*
Data validation	*"the quality and validity of the used data is of course important."*
Data ethics	*"(–) whether it is ethical to use the data, and in what ways."*
Model versioning	*"We need to version each model and be able to connect them to datasets they were trained with."*
Algorithm specifics	*"(–) a lot depends on what kind of a model you are building, is it supervised learning or some reinforcement learning thing."*
Parameter weights	*"After we get one model trained, a lot of time is spent on tweaking the model parameters to achieve best possible performance."*
Manual overview	*"We always also manually follow what the models do and whether the outputs make sense to us."*
Test dataset	*"A ML model cannot be meaningfully tested with anything else than data from the real world situation it's going to be used in."*
Operation environment	*"Of course we need to also test how well the model fits to the surrounding system."*
Model interpretation	*"as complexity grows and there are so many parameters at some point no one is able to understand what is going on."*
XAI tools	*"for example, there's a tool that can deliver a heatmap that explains the factors contributing to the model output [in image recognition]."*
Bias detection	*"detecting biased or unfair outputs of the model... (–)"*
Anomaly detection	*"We can teach a model what is normal and use that in monitoring."*

Materiais e Métodos do Estudo

- **Resultados da análise dos dados:**



Descobertas (Design do Sistema)

- Alguns **pré-requisitos** precisam ser validados antes de se iniciar o SDLC com IA:
  - **Casos de Negócio**
    - Adequação do ML em relação ao caso de negócios.
    - Questões chave definidas. Nem todos os casos de negócio são tratáveis com IA.
    - Necessidades de nível de projeto (desafio com dados confidenciais, regulamentação pesada e forte transparência).

Descobertas (Design do Sistema)

- Alguns **pré-requisitos** precisam ser validados antes de se iniciar o SDLC com IA:
  - **Recursos de Dados**
    - Além da existência, a qualidade e validade dos dados são preocupações da governança de IA.
    - É preciso também haver capacidade, inclusive ética e legal, do uso dos dados.

Descobertas (Design do Sistema)

- Alguns **pré-requisitos** precisam ser validados antes de se iniciar o SDLC com IA:
  - A**mbiente Externo**
    - O ambiente externo, no qual o sistema ML opera, também apresenta requisitos de governança.
    - A governança de IA deve contemplar os requisitos do sistema de TI no qual ele se insere, incluindo requisitos regulatórios, particularmente críticos no que tange a dados.

Descobertas (Desenvolvimento do Sistema)

- A **governança de IA** deve cobrir os diversos processos de desenvolvimento
  - **Desenvolvimento de Dados**
    - É importante versionar os dados de treinamento atrelados aos modelos.
    - A validação de dados é ponto focal da governança, embora nem sempre seja adequadamente tratada.

Descobertas (Desenvolvimento do Sistema)

- A **governança de IA** deve cobrir os diversos processos de desenvolvimento
  - **Desenvolvimento do modelo**
    - O processo é muito experimental, então é necessário controlar as diversas versões de modelos e hiperparâmetros para não perder o controle.
    - A governança de modelo muda de acordo com o tipo de modelo. Por exemplo, modelos supervisionados precisam de controle de qualidade da rotulagem.

Descobertas (Desenvolvimento do Sistema)

- A **governança de IA** deve cobrir os diversos processos de desenvolvimento
  - **Testes do sistema**
    - Governança manual é importante, principalmente quando há muita automação.
    - A boa qualidade dos dados de teste é a principal ferramenta de validação e foco da governança de IA no quesito testes.

Descobertas (Operação do Sistema)

- Existem, dentre outros, 2 temas relevantes para governança de IA na **fase de operação**:
  - **Explicação do Modelo**
    - Modelos de IA de caixa preta são tão complexos que sua compreensão não é humanamente possível.
    - Deve-se considerar os requisitos legais e regulatórios de explicabilidade da solução para criar modelos que os atendam, mesmo a algum custo dos resultados do mesmo.

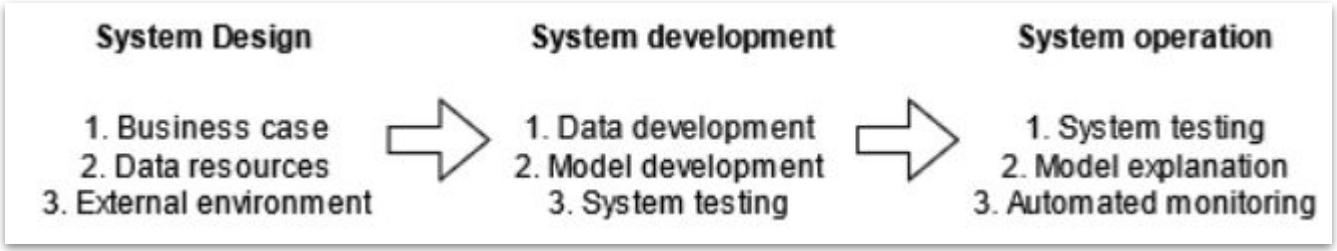
Descobertas (Operação do Sistema)

- Existem, dentre outros, 2 temas relevantes para governança de IA na **fase de operação**:
  - **Monitoramento Automatizado**
    - Detecção de previsões injustas feitas pelo modelo e anomalias / ocorrências e eventos incomuns na saída do sistema de IA ou no sistema geral.
    - Ponto crítico para governança em IA por ser a última linha de defesa.
    - Monitoramento de dados desde a entrada até as saídas dos modelo de ML.

Discussão (Mapeando as Descobertas)

- Com os resultados das entrevistas, foi possível criar uma adaptação do ciclo de vida dos sistemas de IA para o ciclo de vida de sistemas tradicionais de engenharia de software.

Usando um ciclo de desenvolvimento do tipo *waterfall:*

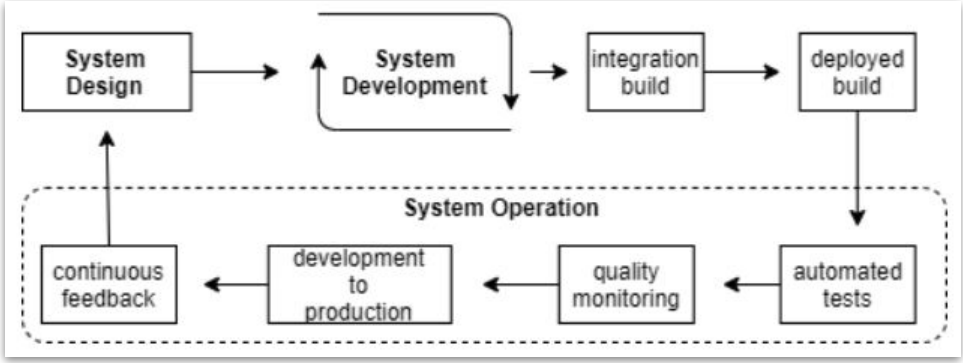


Contras: essa abordagem já caiu em desuso pelos profissionais de software.

Discussão (Mapeando as Descobertas)

- Atualmente são utilizadas técnicas de desenvolvimento baseadas em DevOps, com as práticas de Integração e Entrega Contínuas (CI/CD) sendo um dos maiores pilares.

Usando um ciclo de desenvolvimento com CI/CD:



Discussão (Contribuições)

- Trabalhos anteriores apresentaram um resultado mais **detalhado**, porém também mais **específico** no uso de técnicas e ferramentas.
- Como os sistemas de IA são bastante heterogêneos, uma esteira de CI/CD mais **genérica** é a principal contribuição deste trabalho.
- Ajudar pessoas leigas no assunto a conceitualizar e entender alguns dos problemas específicos de governança em sistemas de IA.

Discussão (Contribuições)

Table 3: A summary of the main theoretical contributions of this study.

Contribution area	Key contributions
AI Governance	We discover key AI governance steps involved in ML model development through expert interviews. We bridge the gap between AI governance theory and practical development. We connect governance into three development stages and show how these connect to a CI/CD pipeline.
SDLC models	We propose a set of governance needs to be taken into account in the SDLC process of systems involving ML. We propose a CI/CD visualization based on [52] and elucidate the AI governance steps involved in each phase of the pipeline.

Discussão

Limitações

- Os entrevistados são todos conhecidos ou estão na *network* dos autores: isso limita a generalização dos dados.
- A novidade do tópico: novas tecnologias e sistemas de IA surgem diariamente.

Trabalhos Futuros

- Resultados podem ser mais detalhados ainda ao apresentá-los de volta aos entrevistados para validação e *feedback*.
- Focar em ferramentas atuais, como o *Amazon SageMaker*, e estudar o suporte à governança de IA.
- Focar, nas entrevistas, na ética em IA.

Conclusões

- Com as entrevistas com especialistas da área, foram descobertos conceitos e temas importantes relacionados à governança de IA
- Com esses conceitos, foi possível adaptar os sistemas de IA para uma *pipeline* de CI/CD.
- Esses passos podem mudar com o avanço da IA e mudanças nas regulamentações e necessidade de maior explicabilidade dos modelos.

Obrigado.