Software Engineering for AI

Jan Bosch

Director Software Center
www.software-center.se
Professor of Software Engineering
Chalmers University of Technology
Gothenburg, Sweden.
www.janbosch.com

Initiative Seminar, Chalmers AI Research, Lindholmen, March 2019
What’s the Difference?

**Characteristics**
- Small-scale
- Few people
- Pilots & prototypes
- No/little value
- No risk

**Characteristics**
- Large scale
- Hundreds/thousands of people
- Industrial, critical deployment
- High value
- High risk

Most Academic Research

Engineering Research
“One of the biggest problems facing businesses is getting AI from pilot phase to scalable deployment.”

KPMG Consulting

“The AI deployment of integrated and scalable solutions across the business requires more than just...”
Five Most Important Jobs in AI

- **AI Architect** – These specialists look at individual business processes - as well as the big picture organization - and determine where they can inject and embed AI successfully. They are also responsible for measuring performance, and sustaining the AI model over time – ensuring it removes mundane tasks to optimize humans in the workforce. The lack of AI architects is a big reason why companies cannot successfully sustain AI initiatives.

- **AI Product Manager** – Working closely with product managers, an AI product manager serves as a liaison across multiple business teams to ensure AI is implemented. They also work closely with these teams – as well as HR – to identify areas that can be automated. This ensures optimal performance of both humans and machines.

- **Data Scientist** – With the ever-growing amount of data available to businesses, there is a shortage of experts with the skills to clean this data, and then design and apply the appropriate algorithms to glean meaningful insights.

- **Software Engineer** – One of the biggest problems facing businesses is getting AI from pilot phase to scalable deployment. Software engineers work hand-in-hand with data scientists to bring AI into production, blending business acumen with a deep understanding of how AI works.

- **AI Ethicist** – As ethical and social implications of AI continue to unfold, companies may need to create new jobs tasked with the critical responsibility of establishing AI frameworks that uphold company standards and codes of ethics. Initially, these roles could be fulfilled by existing leaders in an organization, but as the effects of AI fully take shape, it may need to be the responsibility of one person to ensure these guidelines are upheld.

https://qz.com/work/1517594/the-five-most-important-new-ai-jobs-according-to-kmpg/
Three Key Take-Aways

• **Digitalization** (software, data & AI) is disrupting industry and society to an extent that we have only seen the early beginnings of

• **Productifying** AI/ML/DL solutions and deploying in industrialized contexts is a major engineering challenge

• **Vastly more research** on software engineering for AI (**SE4AI**) is required to address this challenge
Software Center

Mission: Improve the *digitalization* capability of the European Software-Intensive industry with an order of magnitude

Theme: Fast, continuous deployment of customer value

Success: Academic excellence
Success: Industrial impact
Holistic DevOps Framework

**Requirements driven development**
- Regulatory features
- Competitor parity features
- Commodity features

**Outcome/data driven development**
- Value hypothesis
- New "flow" features
- Innovation

**AI driven development**
- Minimize prediction errors
- Many points in data set
- Combinatorial explosion of alternatives

---

Holistic DevOps Framework

System in operation

continuous deployment

behavior data

AI component

SW component

continuous deployment

behavior data
Why Software Engineering For Deep Learning?

How AI Evolves In Industry

- Experimentation & prototyping
- Non-critical deployment of ML/DL components
- Critical deployment of ML/DL components
- Cascading deployment of ML/DL components
- Autonomous ML/DL components

Software Engineering for AI/ML/DL (Operational AI)
Where The Effort Goes ...

- assemble datasets
- create models
- train and evaluate
- deployment

Amount of effort

Amount of attention
Study #1

A. Project EST: Real Estate Valuation
B. Project OIL: Predicting Oil and Gas Recovery Potential
C. Project RET: Predicting User Retention
D. Project WEA: Weather Forecasting
E. Project CCF: Credit Card Fraud Detection
F. Project BOT: Poker Bot Identification
G. Project REC: Media Recommendations
Engineering Challenges for DL Systems

Development Challenges

• **Experiment Management**: how to keep track of all the contextual factors (e.g. HW, platform, etc.) when running many experiments.

• **Limited Transparency**: neural networks give little insight how why and how they work.

• **Troubleshooting**: large libraries, lazy execution and lack of tooling complicate solving defects enormously.

• **Resource Limitations**: large data sets and complex models cause the use of distributed architectures that further complicate experiment management and troubleshooting and increase cost, required knowledge and time.

• **Testing**: testing of data is very complicated and there is little tooling. Also testing models and the underlying infrastructure is very complex.
Production and Organizational Challenges

Production Challenges

• **Dependency Management**: GPU hardware and SW platforms are improving and evolving very rapidly and software often encodes contextual knowledge for performance reasons.

• **Monitoring and Logging**: models are often retrained frequently and behaviour may change. Distinguishing between bugs and improvements is difficult.

• **Unintended Feedback Loops**: real world may adopt to the model, rather than the other way around.

Organizational Challenges

• **Effort Estimation**: we fundamentally don’t know how long it will take to develop an acceptable model (and if it will ever happen).

• **Privacy and Data Safety**: as we often don’t know how neural networks work, it is difficult to guarantee that no personal or private data is stored in the network.

• **Cultural Differences**: data scientists and software engineers often have very different drivers and ways of working.
## Study #2

<table>
<thead>
<tr>
<th>Case</th>
<th>Domain</th>
<th>Use case of ML/DL components</th>
<th>Interviewed experts</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Automotive</td>
<td>Interpreting sensor data to understand the contained information at a high level</td>
<td>P1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(i) Automating tagging of sentiments in online music library (ii) predicting quality of end product based on different measurements from machines, IoT devices of pulp processing and quality measures</td>
<td>P2, P3, P4, P5</td>
</tr>
<tr>
<td>B</td>
<td>Web</td>
<td>Collaborative annotation of training data and predicting quality of annotations</td>
<td>P6, P7</td>
</tr>
<tr>
<td>C</td>
<td>Web</td>
<td>Predicting failures at site to give insights into mobile network operations</td>
<td>P8, P9, P10</td>
</tr>
<tr>
<td>D</td>
<td>Telecom</td>
<td>Automating information extraction from out-of-office reply to optimize communication between sales reps and prospects</td>
<td>P11</td>
</tr>
<tr>
<td>E</td>
<td>Web</td>
<td>Models of ML/DL components of e.g., web search engine, are compared using important measures through A/B tests</td>
<td>P12, P11*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Manager of DL organisation</td>
</tr>
<tr>
<td>P2</td>
<td>Data scientist</td>
</tr>
<tr>
<td>P3</td>
<td>Head of data science team</td>
</tr>
<tr>
<td>P4</td>
<td>Data scientist</td>
</tr>
<tr>
<td>P5</td>
<td>UX lead</td>
</tr>
<tr>
<td>P6</td>
<td>Co-founder</td>
</tr>
<tr>
<td>P7</td>
<td>ML engineer</td>
</tr>
<tr>
<td>P8</td>
<td>Project Tech lead</td>
</tr>
<tr>
<td>P9</td>
<td>Senior researcher</td>
</tr>
<tr>
<td>P10</td>
<td>Researcher</td>
</tr>
<tr>
<td>P11</td>
<td>VP data science</td>
</tr>
<tr>
<td>P12</td>
<td>Data scientist, experimentation</td>
</tr>
<tr>
<td>P11*</td>
<td>Principal data scientist</td>
</tr>
</tbody>
</table>

**Table 1.** Description of domain of studied software-intensive systems, ML/DL components, and roles of interviewees (*previous work*)
## Challenges ML/DL Evolution

<table>
<thead>
<tr>
<th></th>
<th>Experiment Prototyping</th>
<th>Non-critical deployment</th>
<th>Critical deployment</th>
<th>Cascading deployment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>assemble dataset</strong></td>
<td>Issues with problem formulation and specifying desired outcome</td>
<td>Data silos, scarcity of labelled data, imbalanced training set</td>
<td>Limitations in techniques for gathering training data from large-scale, non-stationary data streams</td>
<td>Complex and effects of data dependencies</td>
</tr>
<tr>
<td><strong>create model</strong></td>
<td>Use of non-representative dataset, data drifts</td>
<td>No critical analysis of training data</td>
<td>Difficulties in building highly scalable ML/DL pipeline</td>
<td>Entanglements causing difficulties in isolating improvements</td>
</tr>
<tr>
<td><strong>train and evaluate model</strong></td>
<td>Lack of well-established ground truth</td>
<td>No evaluation of models with business-centric measures</td>
<td>Difficulties in reproducing models, results and debugging DL models</td>
<td>Need of techniques for sliced analysis in final model</td>
</tr>
<tr>
<td><strong>deploy model</strong></td>
<td>No deployment mechanism</td>
<td>Training-serving skew</td>
<td>Adhering to stringent serving requirements e.g., of latency, throughput</td>
<td>Hidden feedback-loops and undeclared consumers of the models</td>
</tr>
</tbody>
</table>

Table 2. Challenges in the evolution of use of ML/DL components in software-intensive systems
Conclusions

• **Digitalization** (software, data & AI) is disrupting industry and society to an extent that we have only seen the early beginnings of

• **Productifying** AI/ML/DL solutions and deploying in industrialized contexts is a major engineering challenge

• **Vastly more research** on software engineering for AI (**SE4AI**) is required to address this challenge