Meditations on First Deployment

A Practical Guide to Responsible Development

EuroPython 2020

Alejandro Saucedo

@AxSaucedo
Hello, my name is Alejandro

Alejandro Saucedo
@AxSaucedo
The magic of programming

You can wake up with an idea
and have a prototype by the end of day/weekend.
Software is eating the world

The future wonders of the world will be running Python
Critical infrastructure increasingly depends on running software

...and regardless of the software / hardware abstractions, the impact will always be human, at an individual and societal level
Urgency vs Best Practice
AND

Cybersecurity Attacks

Software Outages

Misuse of personal data

Algorithmic Bias

The impact of a bad solution can be worse than no solution at all.
Responsibility Infrastructure

Team / Delivery Process
- Cross functional skillset
- Key domain experts
- Accountability structure
- Principled alignment
- Relevant delivery structure

Individual Practitioner
- Technology best practices
- Most relevant tools
- Competence in field
- Professional responsibility

Department / Organisation
- High level Principles
- Governing structure
- Aligned objectives
- Escalation structure
As software developers we have a growing professional responsibility to our craft.

<table>
<thead>
<tr>
<th></th>
<th>Empowered</th>
<th>Unempowered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethical</td>
<td>😎</td>
<td>😞</td>
</tr>
<tr>
<td>Unethical</td>
<td>😈</td>
<td>😄</td>
</tr>
</tbody>
</table>

- Ethical
- Empowered
- Ought to do good
- Know how to
Large ethical challenges cannot fall on the shoulders of a single software developer.
End-to-end Approach

1. **Principles & Guidelines**
   High level guidelines that provide a principled approach towards designing, building and operating machine learning.

2. **Industry standards & regulatory frameworks**
   Practical guidelines that set the bar for requirements around risk assessment and evaluation for machine learning systems.

3. **Open Source Software**
   Practical implementations of the best practices on the infrastructure that provides the backbone to most applications.
**Terminology**

**Ethics**
Moral principles that govern a person's behaviour or the conducting of an activity.

**Principles**
Fundamental truths or propositions that serve as the foundation for a system of belief or behaviour or for a chain of reasoning.

**Why not just follow existing rules?**
When dealing with new technologies/situations, there may just not be enough examples to base on, but practitioners will need to make decisions.
Whose Ethics?

Eastern? Western? ...

Philosophical Foundations

Current (Geo)political ecosystem

Understanding underlying philosophical foundations allows us to understand where we come from, to come to more powerful mutual agreements.
Principles & Ethics Framework

The ACM’s Code of Ethics & Professional Conduct

<table>
<thead>
<tr>
<th>Principle</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Human Augmentation</td>
<td>I commit to assessing the impact of incorrect predictions and when reasonable, design systems with backups of the basic decision process.</td>
</tr>
<tr>
<td>2. Bias Evaluation</td>
<td>I commit to continuously develop processes that allow me to understand, document, and monitor risks in development and production.</td>
</tr>
<tr>
<td>3. Explainability by design</td>
<td>I commit to develop tools and processes to continuously improve transparency and explainability of models and learnings available when responsible.</td>
</tr>
<tr>
<td>4. Reproducible systems</td>
<td>I commit to continuously improve my machine learning infrastructure for a responsible model reproducibility.</td>
</tr>
<tr>
<td>5. Objectivity</td>
<td>I commit to continuously improve my machine learning infrastructure for a responsible model reproducibility.</td>
</tr>
<tr>
<td>6. Practical Accuracy</td>
<td>I commit to develop processes to ensure my accuracy and cost are in line with the seen risks and values attached to the ethical aspects of applications.</td>
</tr>
<tr>
<td>7. Trust beyond the user</td>
<td>I commit to build and communicate processes that protect and handle data with stakeholders that may differ from the teams directly involved in it.</td>
</tr>
<tr>
<td>8. Data-risk awareness</td>
<td>I commit to developing and using reasonable processes and infrastructure to protect our data and model security while being open to consultation during the development of machine learning systems.</td>
</tr>
</tbody>
</table>
Principles = good for business and software!

<table>
<thead>
<tr>
<th>Contribute to society and to human well-being...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avoid harm</td>
</tr>
<tr>
<td>Be honest and trustworthy</td>
</tr>
<tr>
<td>Be fair and take action not to discriminate</td>
</tr>
<tr>
<td>Respect the work required to produce new ideas...</td>
</tr>
<tr>
<td>Respect privacy</td>
</tr>
<tr>
<td>Honor confidentiality</td>
</tr>
</tbody>
</table>

| Strive to achieve high quality...               |
| Maintain high standards...                      |
| Know and respect existing rules...             |
| Accept and provide appropriate professional review|
| Perform work only in areas of competence       |
| Foster public awareness and understanding...    |
| Access computing and communication resources only when authorized |
| Design and implement systems that are robustly and usably secure |
Industry/Code Standards

Who sets code/industry standards? You!

Who uses the industry standards? Maybe You! and maybe them too...

Standard: A repeatable, harmonised, agreed & documented way of doing something
Standardisation Bodies

You can get involved in the design and development and use of standards
Open Source as Foundation

Open source is now becoming the backbone for critical infrastructure that runs our society.
Open Source as Policy

Principles are useless if the foundation is not in place to introduce and manage

1. Principles & Guidelines
   High level guidelines that provide a principled approach towards designing, building and operating machine learning.

2. Open Source Software
   Practical implementations of the best practices on the infrastructure that provides the backbone to most applications.
Open Source as Lead

Open source leaders are developing the core cogs that regulation depends on.

1. **Principles & Guidelines**: High level guidelines that provide a principled approach towards designing, building and operating machine learning.

2. **Open Source Software**: Practical implementations of the best practices on the infrastructure that provides the backbone to most applications.
Open Source Foundations

You can get involved on the design and development and use of standards
Sidenote: Regulation

We all can agree: Bad regulation is BAD.

However good regulation can be a catalyst for innovation through enforcement of best practices and mitigation of bad actors.
Software’s Massive Traction

- Internet Services
- Machine Learning Automation
- Cloud Native infrastructure
- Gaming and design tools
- Etc, etc, etc, etc

Growth
Not all can be solved w code

Problems in the world

When you run around with a hammer, everything may look like a nail
The Challenge of our Generation

Societal Impact

Economic Impact
And potentially not the last

Ensuring the right solution

Before tackling a problem we should be able to identify how much of it is actually a software problem before actually writing code.

And whether the solution is even solving a problem.
Practical Deep Dive
Production machine learning systems
Prod ML Systems are HARD

Specialised Hardware (GPU, etc)

Complex Dependency Graphs

Compliance

Reproducibility of components

Last year’s talk on the challenges & landscape in ML: https://www.youtube.com/watch?v=Ynb6X0KZKxY
Principles for responsible AI

1. Human augmentation / review
2. Bias evaluation capabilities
3. Explainability by justification
4. Reproducible ops infrastructure
5. Displacement strategy
6. Practical statistical metrics
7. Trust by privacy
8. Security risks

http://ethical.institute/principles.html
Procurement Framework

A set of templates for industry practitioners:

- Request for proposal
- ML maturity model
- Tender competition template

http://ethical.institute/rfx.html
ML Maturity Model

From principles to a checklist

- Each has a set of questions for supplier compliance
- Top-bottom approach providing red flags

<table>
<thead>
<tr>
<th>Practical benchmarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explainability by justification</td>
</tr>
<tr>
<td>Infrastructure for reproducible operations</td>
</tr>
<tr>
<td>Data and model assessment processes</td>
</tr>
<tr>
<td>Privacy enforcing infrastructure</td>
</tr>
<tr>
<td>Operational process design</td>
</tr>
<tr>
<td>Change management capabilities</td>
</tr>
<tr>
<td>Security risk processes</td>
</tr>
</tbody>
</table>

http://ethical.institute/rfx.html
# Alignment on first principles

<table>
<thead>
<tr>
<th>#</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>Supplier doesn’t have infrastructure and/or processes to version different machine learning models where reasonable</td>
</tr>
<tr>
<td>#2</td>
<td>Supplier does not have a protocol to evaluate whether new ML model requires domain expert for evaluation of low confidence results</td>
</tr>
<tr>
<td>#3</td>
<td>Supplier system doesn’t have capabilities to perform development across production and QA/BETA environments</td>
</tr>
<tr>
<td>#4</td>
<td>Supplier does not have a process and/or infrastructure to revert models in production without unreasonable level of disruption</td>
</tr>
<tr>
<td>#5</td>
<td>Supplier doesn’t have processes and/or infrastructure that ensures only users with explicitly granted permissions have access to PII data</td>
</tr>
<tr>
<td>#6</td>
<td>Supplier doesn’t have process to assess human review process requirements based on the impact of incorrect predictions</td>
</tr>
<tr>
<td>#7</td>
<td>No process and/or infrastructure to ensure machine learning data encrypted on transport/rest</td>
</tr>
<tr>
<td>#8</td>
<td>Supplier doesn’t have a process and/or infrastructure to introduce specialised model evaluation metrics where required</td>
</tr>
</tbody>
</table>

http://ethical.institute/rfx.html
Broader list of Prod OSS libraries

Broader list of guidelines

[GitHub repository link]

https://github.com/EthicalML/awesome-artificial-intelligence-guidelines
<table>
<thead>
<tr>
<th>CORE</th>
<th>SECONDARY</th>
</tr>
</thead>
<tbody>
<tr>
<td>#2 Bias evaluation</td>
<td>#8 Security</td>
</tr>
<tr>
<td>#3 Explainability</td>
<td>#1 Human-in-the-loop</td>
</tr>
<tr>
<td></td>
<td>#6 Practical metrics</td>
</tr>
</tbody>
</table>
Loan approval process

Domain expert evaluates application
Loan is approved or rejected
Manual process

Business wants to automate this process with machine learning
Traditional data science process
We obtain some data

<table>
<thead>
<tr>
<th>age</th>
<th>workclass</th>
<th>education</th>
<th>education-num</th>
<th>marital-status</th>
<th>occupation</th>
<th>relationship</th>
<th>ethnicity</th>
<th>gender</th>
<th>capital-gain</th>
<th>capital-loss</th>
<th>hours-per-week</th>
<th>native-country</th>
<th>loan</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>State-gov</td>
<td>Bachelors</td>
<td>13</td>
<td>Never-married</td>
<td>Adm-clerical</td>
<td>Not-in-family</td>
<td>White</td>
<td>Male</td>
<td>2174</td>
<td>0</td>
<td>40</td>
<td>United-States</td>
<td>False</td>
</tr>
<tr>
<td>50</td>
<td>Self-emp-not-inc</td>
<td>Bachelors</td>
<td>13</td>
<td>Married-civ-spouse</td>
<td>Exec-managerial</td>
<td>Husband</td>
<td>White</td>
<td>Male</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>United-States</td>
<td>False</td>
</tr>
<tr>
<td>38</td>
<td>Private</td>
<td>HS-grad</td>
<td>9</td>
<td>Divorced</td>
<td>Handlers-cleaners</td>
<td>Not-in-family</td>
<td>White</td>
<td>Male</td>
<td>0</td>
<td>0</td>
<td>40</td>
<td>United-States</td>
<td>False</td>
</tr>
<tr>
<td>53</td>
<td>Private</td>
<td>11th</td>
<td>7</td>
<td>Married-civ-spouse</td>
<td>Handlers-cleaners</td>
<td>Husband</td>
<td>Black</td>
<td>Male</td>
<td>0</td>
<td>0</td>
<td>40</td>
<td>United-States</td>
<td>False</td>
</tr>
<tr>
<td>28</td>
<td>Private</td>
<td>Bachelors</td>
<td>13</td>
<td>Married-civ-spouse</td>
<td>Prof-specialty</td>
<td>Wife</td>
<td>Black</td>
<td>Female</td>
<td>0</td>
<td>0</td>
<td>40</td>
<td>Cuba</td>
<td>False</td>
</tr>
</tbody>
</table>

We get 8000 rows with target
We train our model

99% Accuracy

Time for production?
It’s a disaster
When we look at our data...

Training data

Production data
Let’s analyse dataset further
Bias Evaluation Process

Machine Learning Model Development

1. Get data & clean based on knowledge
2. Business understanding on data bias through risk assessment
3. Define some features to transform data
4. Select model and do parameter search
5. Define scoring metrics and calculate accuracy
6. Business and model/feature/metrics understanding and risk management
7. Persist model & code
8. Gather more data
9. Change current feature/model/accuracy
10. Rinse & Repeat

Prediction with trained model

1. Unseen data ingestion
2. Prediction using feature transformations and trained model
3. Analysis of requirements for infrastructure required for monitoring, as well as process design plans to add human in the loop design
4. Results
We can upsample/downsample...
Taking into account correlation
Much better...
Let’s explain predictions
Let’s explain predictions
We can add manual review
Recap

The impact of software development
Responsibility as individual and organisations
Ethics and Principles
Industry & Code Standards
Finding the right solution for the right problem
Practical deep dive on AI
Meditations on First Deployment
A Practical Guide to Responsible Development

EuroPython 2020
Alejandro Saucedo

@AxSaucedo
Massive Shoutout to what-if.XKCD.com

For their always-amazing artwork & content! Check it out and support them!