PAUL SCHERRER INSTITUT PSI ETH zürich EPFL

Swiss Data Science Center

The Swiss Data Science Center

Every day, transforming your data into knowledge

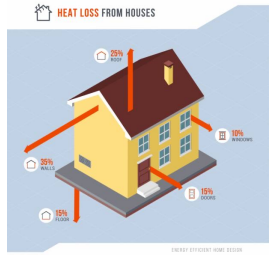
Roberto Castello PhD

Principal Data Scientist - SDSC Innovation team

A few words about me

Data science and Big Data

Renewable energies



Smart buildings

- PhD in Physics from University of Torino, Italy
- Research fellow and data analyst in HEP at CERN
- Senior scientist at EPFL Lab of Solar Energy and Building Physics (LESO-PB)
- Principal Data Scientist at the EPFL Swiss Data Science Center (SDSC)

The Swiss Data Science Center (SDSC)

Swiss Data Science Center

Our mission is to accelerate the adoption of Data Science and AI in Switzerland

An initiative from the ETH Domain

Large multidisciplinary team of 100+ data science professionals

Research, Innovation, Software & platforms, Education

The SDSC ecosystem

Swiss Data Science Center

Academic team (37 ppl)

Helping researchers of EPFL ETHZ and PSI leverage the state-of-the-art in data science

Finance + System eng. + Comm + Admin (13 ppl)

Executive director: Dr. Olivier Verscheure

SDSC

Innovation team (24 ppl)

- Helping organizations advance in their journey to data-driven decision-making
- Implementing advanced AI techniques for industrial applications
- Bringing together stakeholders from different industrial backgrounds and organizing the dialog on business applications of data science

RENKU team (18 ppl)

Reproducible and collaborative data analysis projects

https://renkulab.io

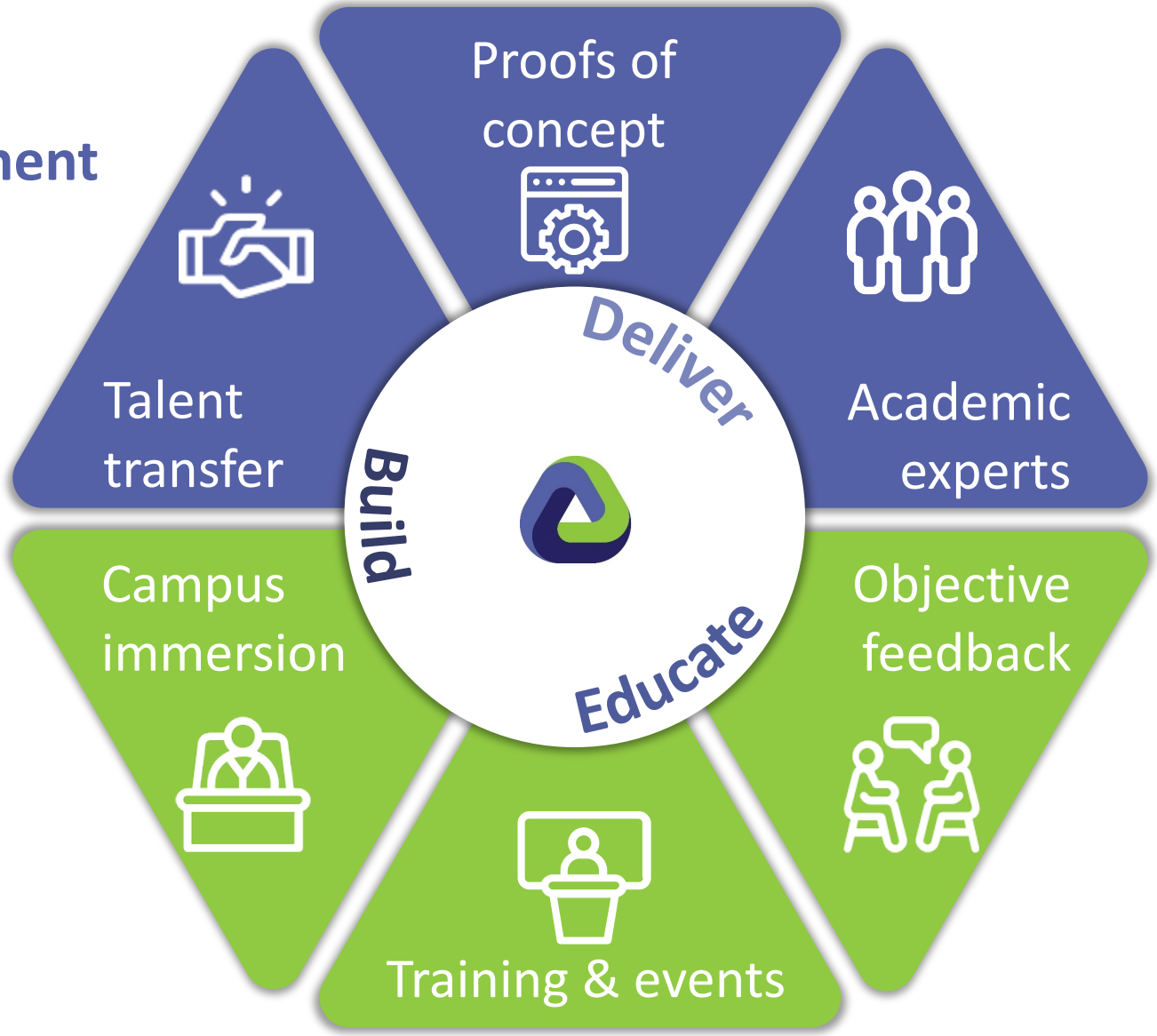
Open Research Data Engagement & Services (12 ppl)

Data infrastructure, security, compliance and FAIR principles

How to collaborate with the SDSC Innovation team

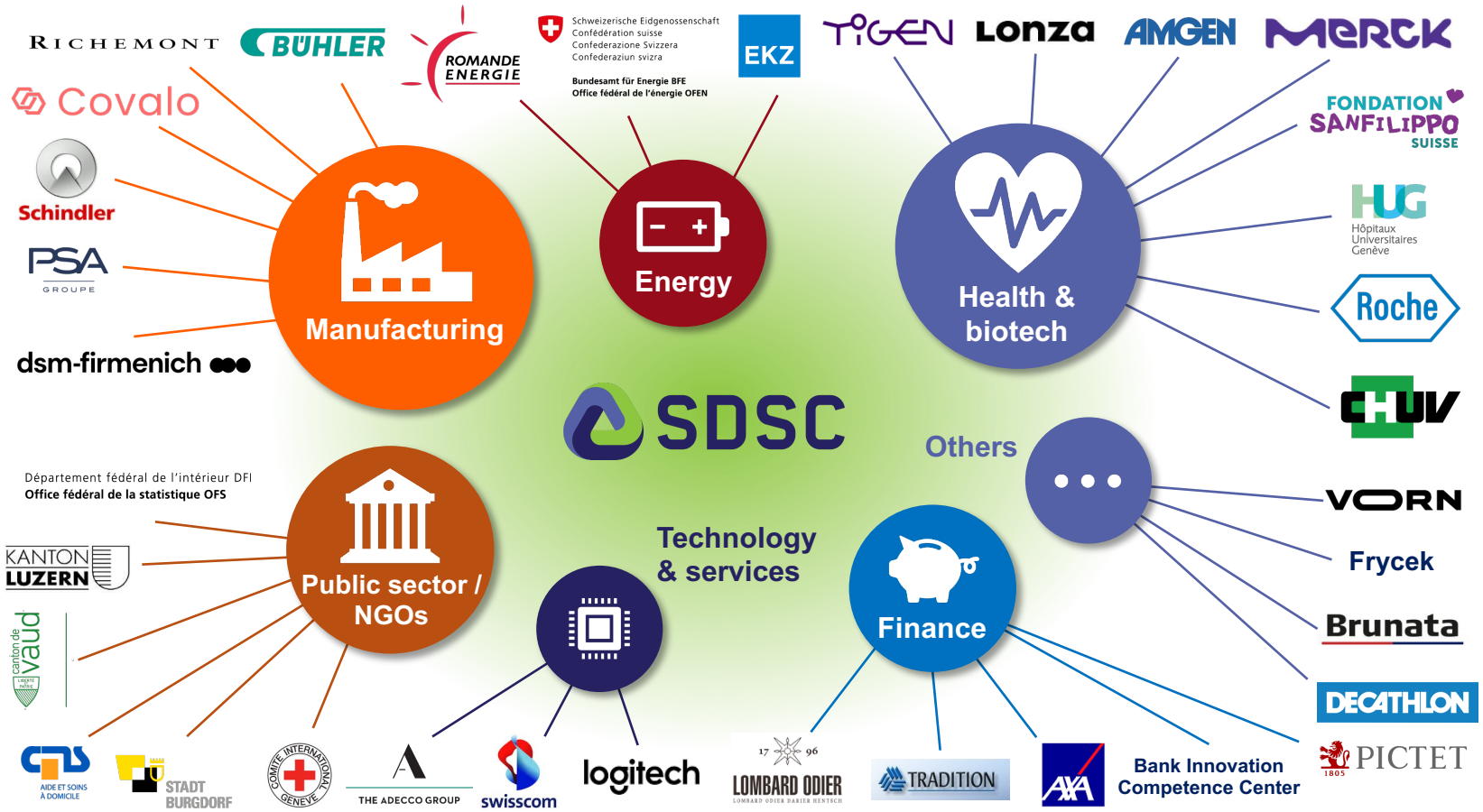
Swiss Data Science Center

Research agreement



Membership

Collaboration landscape

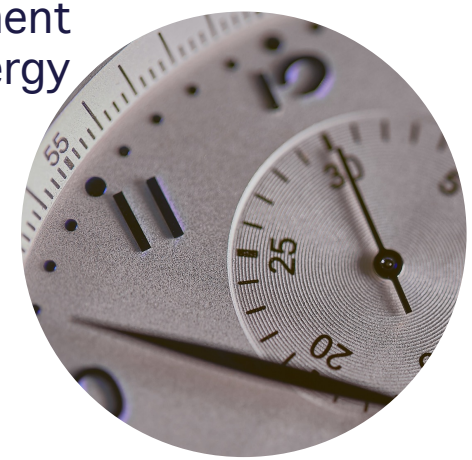


A typical day at the SDSC Innovation team, helping…

Manufacturing: Optimize milling plant component settings to save energy

Job recruiters: Match job offers to candidates based on skills

Private bank: Detect money laundering activities



Retail: Allocate luxury products to market based on demand

Pharma: Predict adverse effects of tumor drugs

...as well Swiss institutions and NGOs

BFE/SFOE: Identify data science potential in the energy domain

BFS/FSO: Expand local data science capabilities in public sector projects

Hospitals: Collaborate on data-driven projects featuring multi-disciplinary teams

Cantons: Support data science initiatives with external expertise

ICRC: Harness the power of AI to uncover patterns of violence

AI for a humanitarian research project

ICRC

CONTEXT

The ICRC wants to **evaluate the impact of its actions** on the evolution of violence. However, it does not have the resource to analyze all events of interests.

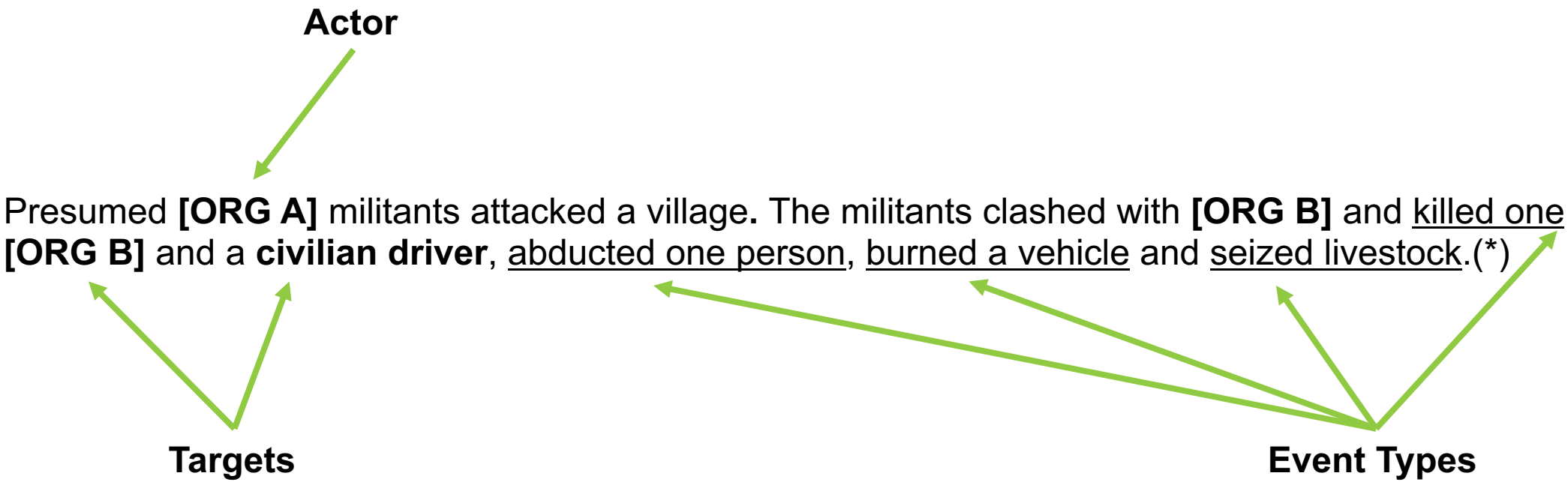
OBJECTIVES

Develop a low resource approach that enables the **automatic classification of violent events** (from free text) into defined event types. The approach should also be able to **identify the actor and target** of violent events.

BENEFITS

Our method can **detect trends in violent events** relying only on pre-trained models and a **limited annotation step**.
From the results, we can retrospectively evaluate the impact of the ICRC work in engaging in confidential talks with the involved armed forces

Classifying unstructured data, automatically



(*) Example of a free text **from the ACLED database, The Armed Conflict Location & Event Data Project:** a disaggregated data collection, analysis, and crisis mapping project built by an NGO https://acleddata.com/#/dashboard

Conventional pipeline

Swiss Data Science Center

Domain experts
design codebook

event type: injury
description
examples

Human annotators
collect and label
training data

injury
abuse
injury

Engineers build and
train classifier

event →  → injury

Our approach in a nutshell

Swiss Data Science Center

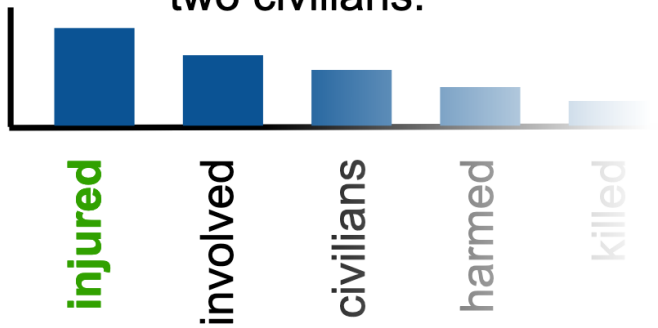
Domain experts design codebook in the form of prompt template and answer candidates directly with engineers

① **Prompting**

event descript **template**

Military injured two civilians. People were [Z]

$[Z^K]$ answer candid.



② **Entailment**

premise Military injured two civilians.

validate hypothesis People were [Z* = **injured**]

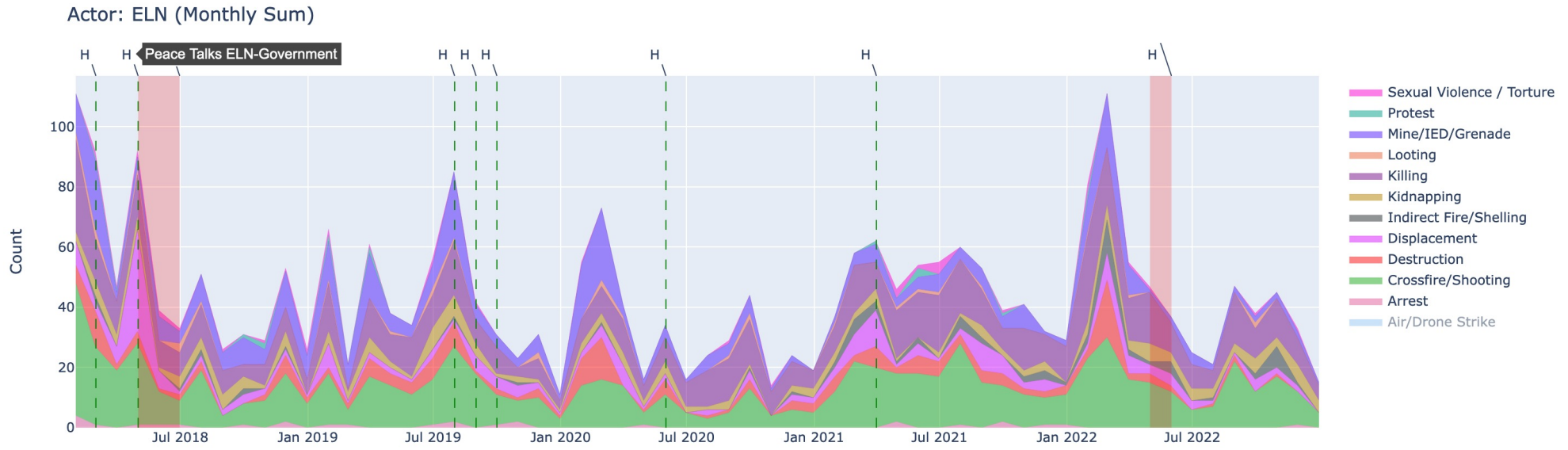
3) Codebook for classification

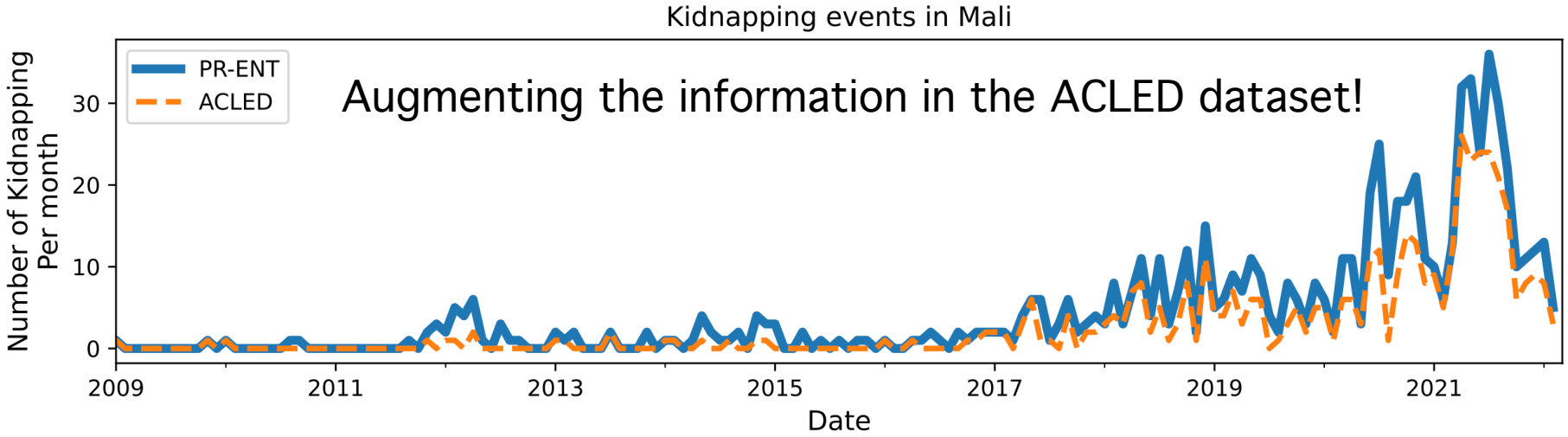
Event Type	Template	Entailed Answer Candidate
Arrest	People were [Z].	arrested AND NOT kidnapped
Killing	This event involves [Z].	killing
	People were [Z].	killed
Looting	This event involves [Z].	looting OR theft OR robbery
Sexual Violence	This event involves [Z].	rape
	People were [Z].	abused OR raped
Kidnapping	This event involves [Z].	kidnapping
	People were [Z].	kidnapped OR abducted
Protest	This event involves [Z].	protest OR demonstration
	People were [Z].	protesting

Examples

Text	Judgments	Hypothesis
A man inspects the uniform of a figure in some East Asian country.	contradiction	The man is sleeping
A soccer game with multiple males playing.	entailment	Some men are playing a sport

Monitoring and augmenting





Swiss Data Science Center

Thank you!

www.datascience.ch

contact@datascience.ch

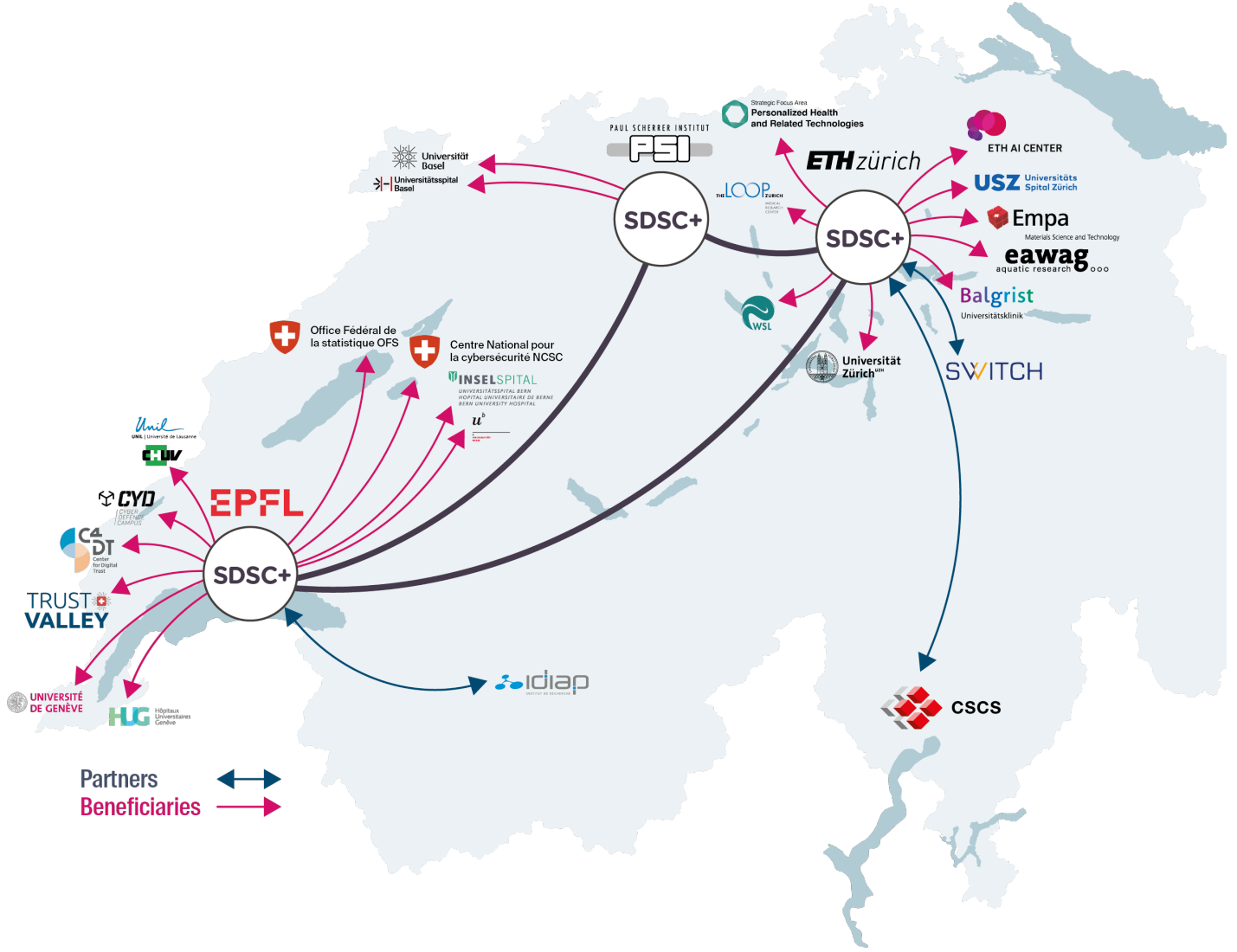
Core mission of the SDSC Innovation team

- **Capability enhancement**: Collaborations result in tangible impact & confidence to invest in AI
- **Capacity building**: Growth of partner's internal talent pool (talent transfer from SDSC) & ability to deliver
- **Cultural shift**: Awareness resulting in transformational choices (systematic data-driven decision making)

A look into the future: SDSC+ (2025 and beyond)

Swiss Data Science Center

- Becoming **a national research infrastructure in the ETH domain** for AI, ML, and data sciences
- Serving universities, industry university hospitals, public institutions, and civil society



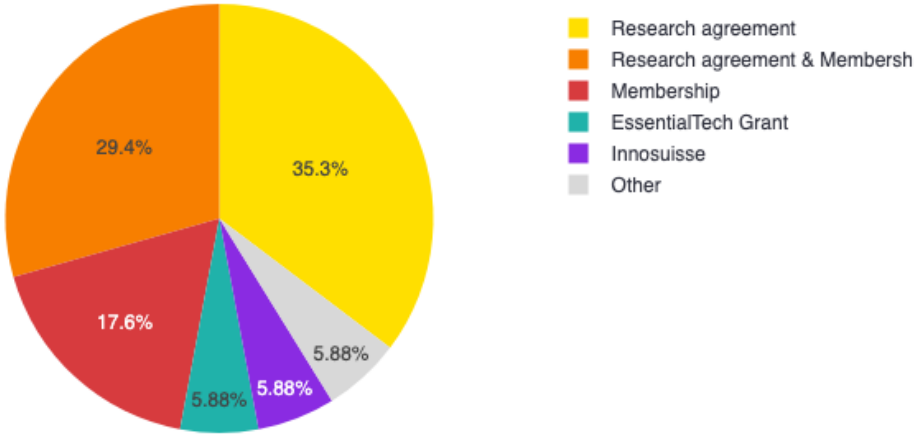
The next SDSC connect event – save the date!

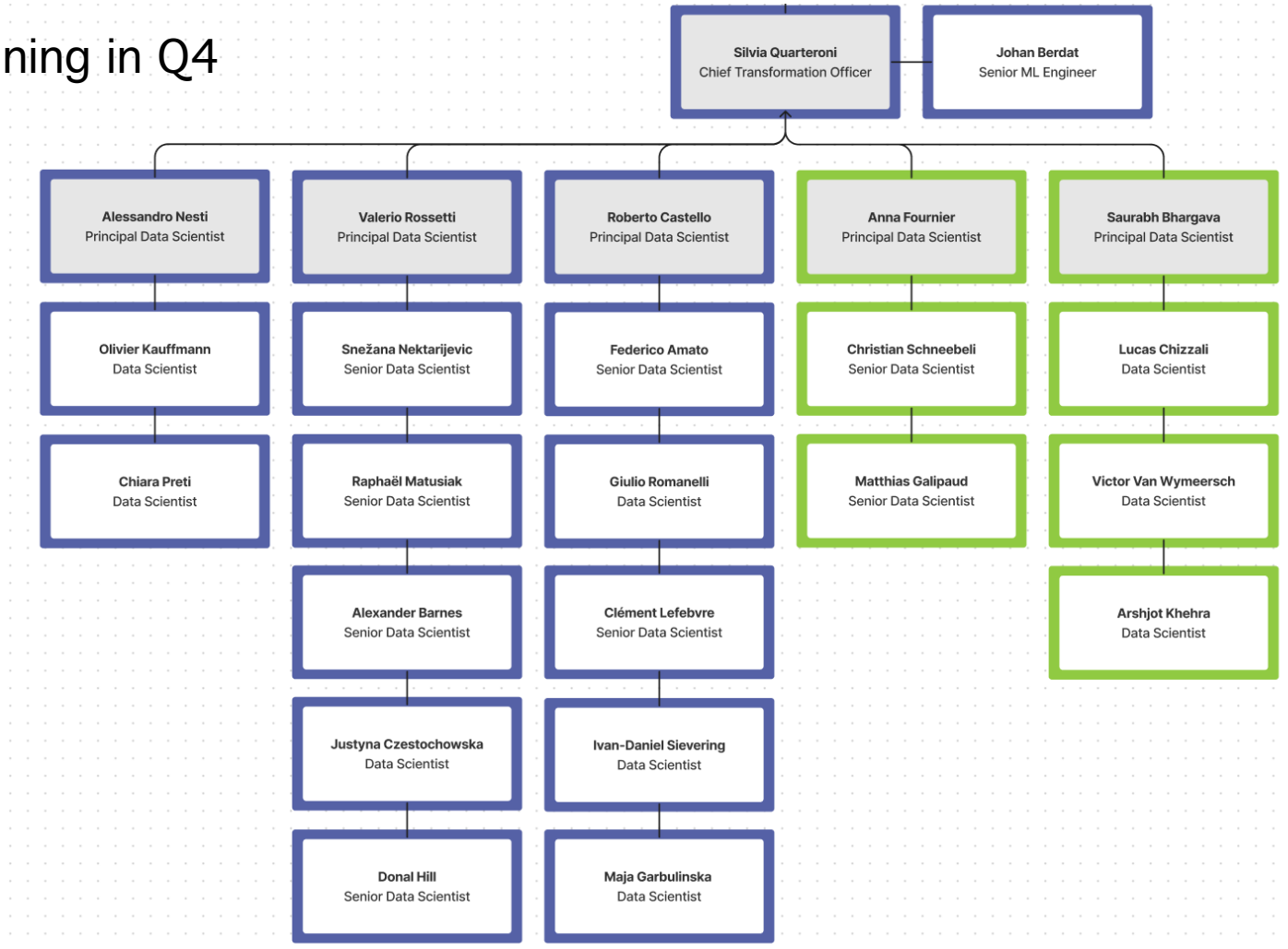
Swiss Data Science Center

Confirmed speakers: Axpo, EKZ, Romande Energie, ETHZ ESC, SFOE, Swissgrid,....

The SDSC innovation team: figures and numbers

- 24 people (6F + 18M) + 2 new people joining in Q4
  - 10 nationalities, average age 32.5 yrs old
  - 9 DS, 9 senior DS/ML engineers, 5 Principal DS
- 20+ collaborations ongoing
- 2 main collaboration models:
  - Research agreement
  - Membership





Innovation collaborates with industry and public sector

Swiss Data Science Center

"*In our journey towards Augmented Creation, SDSC is a key partner, providing strong data science & AI capabilities. Their unique set-up, combining academic and industry expertise, bring true value to our digital strategy.*"

Philippe Glénat, VP Digital Lab, Firmenich

"*After only a few weeks, SDSC was already able to help us demystify important business issues, translate insights into actions and generate true value-added services.*"

Geoffroy Lefebvre, Group Digital Distribution Director, Richemont

"*The collaboration with SDSC has been key to help us unlock the full potential of our data, bringing the relevant expertise at the right time to our different challenges.*"

Quentin Le Masne Director, Connected Health & Devices, Merck Group

"*Our seamless collaboration with the SDSC has helped us to explore the usefulness and applicability of latest, cutting-edge data science methods for our work at SFOE.*"

Fabian Heymann, Digital Innovation Specialist at the Swiss Federal Office of Energy

Forecasting the end-user Swiss electricity consumption

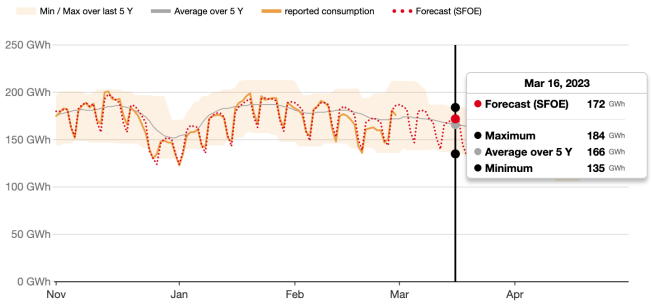
Swiss Data Science Center

CONTEXT

The security of supply of electricity and gas has required dramatic measures in the winter 2022/2023. Swiss policy makers were in urgent **need of a near to real-time estimates and a forecast of the national electricity consumption**. At present, only the vertical load seen from the transmission network and the total national electricity consumption are published **and with major delays**

OBJECTIVES

We use a Generalized Additive Model (GAM) that, by learning from the Swiss historic national consumption, **can predict the national electricity demand using as input also calendar and meteorological data**.



BENEFITS

The model has been **integrated into the Federal Energy Dashboard operated by SFOE** with the objective of monitoring the real-time electricity consumption and quantifying the energy saving in response to the Federal campaign launched in August 2022

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Bundesamt für Energie BFE
Office fédéral de l'énergie OFEN

Quantifying electricity savings per sector

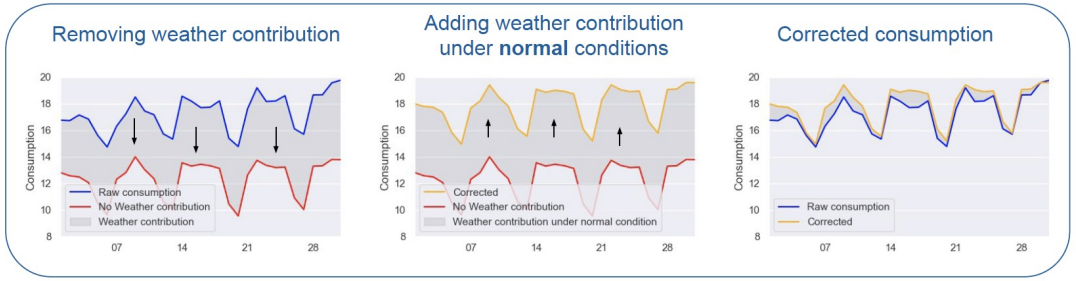
Swiss Data Science Center

CONTEXT

In Sep 2022 the Swiss government has launched an energy saving campaign, **including a request to quantify the change in electricity consumption habits (winter 2022 target: 10% reduction)**

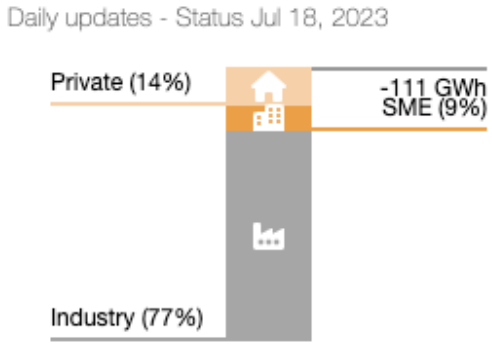
OBJECTIVES

We used a methodology to quantify changes in electricity consumption with a **bottom-up** approach. From **individual load curves from smart-meters of Distribution Network Operators (DNOs) we extrapolate the load to the national level** using scaling factors and corrections



BENEFITS

The model has been **integrated into the Federal Energy Dashboard operated by SFOE** to monitor the daily weather-adjusted over/underconsumption by customer group



Predicting protein properties from primary structure

Swiss Data Science Center

CONTEXT

Biomedical data is both **expensive and time-consuming** to obtain. A pharmaceutical company wanted to **leverage the small amount of** collected **data** to have better predictions of protein properties.

OBJECTIVES

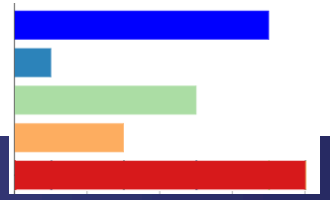
Create a deep learning framework that is able to **accurately predict desired property of a protein** based on its amino acid sequence. We aim to initially train the model on a large set of publicly available data and **fine-tune it with internal dataset**.

BENEFITS

We delivered a flexible framework that captures the underlying **features of ~50M proteins** and has **on-demand fine-tuning** option to accurately predict the property of interest.

Input:
Protein 1: MPRTMLTDQHWK…LSIHHNSNLRNFIEYRI
Protein 2: MIILIGSQKGGGKSI…AINIAAYLAKHKK
Protein 3: MKKPTHKIYRTTNW…AYNRLMSRGNIAIW
Protein 4: MPRFDPATQWYAS…KGKQGRNQT
Protein 5: MYSDAAIQCCLMIK...LFRLSLRMVTGFVQIKLCGL

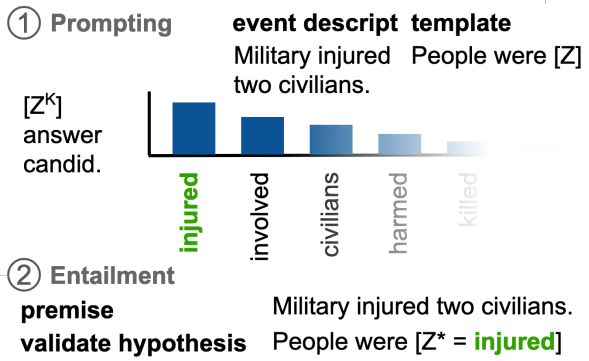
Output:
Estimated properties



Monitoring Patterns of Violence

CONTEXT

The ICRC wants to **evaluate the impact of its actions** on the evolution of violence. However, it does not have the resource to analyze all events of interests.

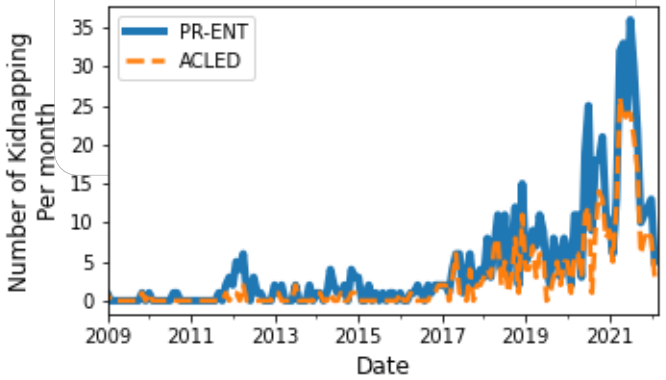


OBJECTIVES

Develop a low resource approach that enables the **automatic classification of violent events** (from free text) into defined event types. The approach should also be able to **identify the actor and target** of violent events.

BENEFITS

Our method can **detect trends in violent events** relying only on pre-trained models and a **limited annotation step**.
It can also detect trends in violent group behavior both in type of violence or target of violence.



SDSC

COMITE INTERNATIONAL GENEVE

Your collaboration opportunities

Swiss Data Science Center

Research agreements

- One or more data scientist(s) hosted and managed by the Swiss Data Science Center
- Hands-on data science work by a full-time SDSC professional
- Project-related meetings with the collaboration lead
- Organization of relevant *use case* workshops
- Ad hoc access to machine learning experts

Memberships & Community building

- **Bespoke coaching/supervision, training & joint thematic workshops**
- Campus immersion for your data scientists
- Support to data science footprint growth & talent retention
- Premium: Support to hiring for Member's AI team

Included in all collaborations

- Invitation to cross-industry events (for data scientists)
- Participation to SDSC Executive Meetups (for executives)
- Connection with academia (e.g. MSc students), social media promotion