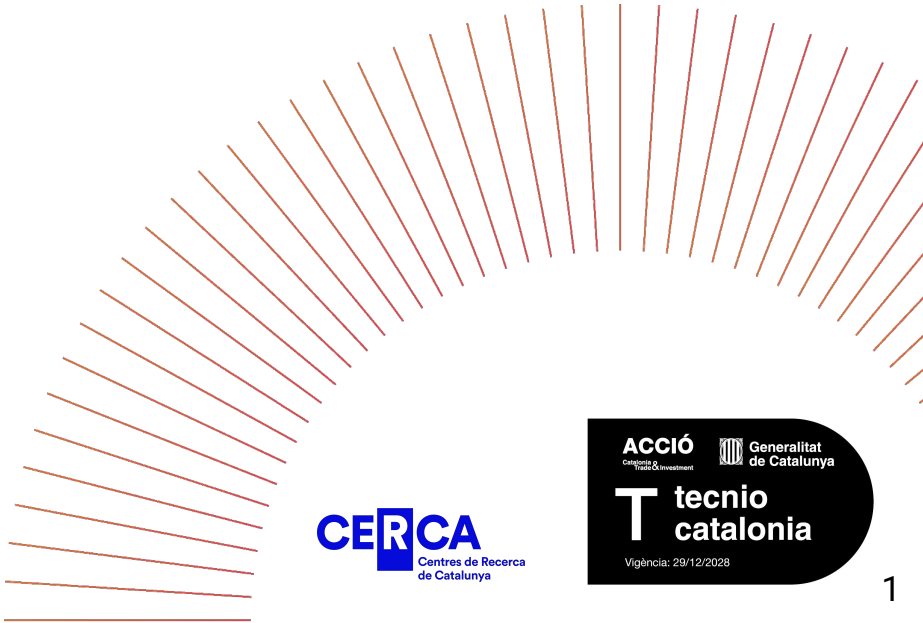
i2cat

The Internet Research Centre

LARA

Latent Representation for Adaptive tasks

SIG-AI Meeting
Albert Calvo, Ph.D.
11/03/2026



CERCA Centres de Recerca de Catalunya

ACCIÓ Catalonia Trade & Investment · Generalitat de Catalunya · tecnio catalonia
Vigència: 29/12/2028

whoami

Albert Calvo, Ph.D.

- Senior Researcher @ i2CAT Foundation
- Trustworthy-AI and Efficient Multimodal AI
- Involvement in different competitive projects
- Research Interests: XAI, SLM/LLM, Confidential-AI

PALANTIR
Practical Autonomous Cyberhealth for resilient SMEs & Microenterprises

ExtremeXP

FIRST

eCrime2023
Symposium on Electronic Crime Research

SECANT

cipher
Cybersecurity Intelligence, Protection, and Holistic Enterprise Resilience

tnc22
NAVIGATING THE UNEXPLORED
Trieste, 13-17 June 2022

tnc23
Tirana, Albania | 5-9 JUNE 2023

i2cat

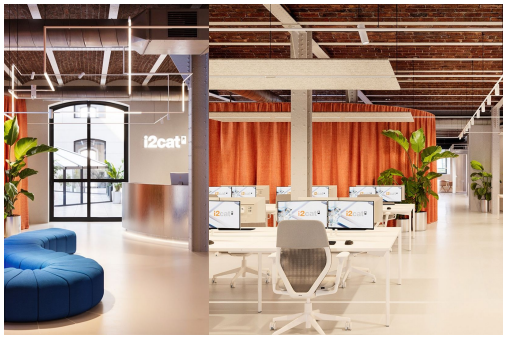
The i2CAT is a non-profit foundation that conducts research and innovation on advanced digital technologies.

5G, Cybersecurity, IoT, VR, AI, blockchain, and space communication

119 R&D projects

74 Total competitive projects

250+ Staff



Geant Innovation Program 2026

- Innovation projects that create reusable POCs for the Geant Ecosystem
- Execution (January to June 2026)
- 6 Funded project (300k in total)

G.O.READY – Feasibility Study for a GÉANT Open Source Program Office (OSPO). SWITCH (🇨🇭) HEANET (🇮🇪)

LARA – LAtent Representation for Adaptive tasks. i2cat (🇪🇸)

FeduMEET – Delivering eduMEET as a Service Using Federated NREN Infrastructures. PCSS (🇵🇱)

CHAMELEON-REN: Stimulus-Driven, Dockerised OWASP Web Application Honeypots for the Global Research and Education Community. The Open University & OWASP (🇬🇧)

EXPATS – Explainable Autofill for Trustworthy Surveys. PXL University of Applied Sciences and Arts (🇧🇪)

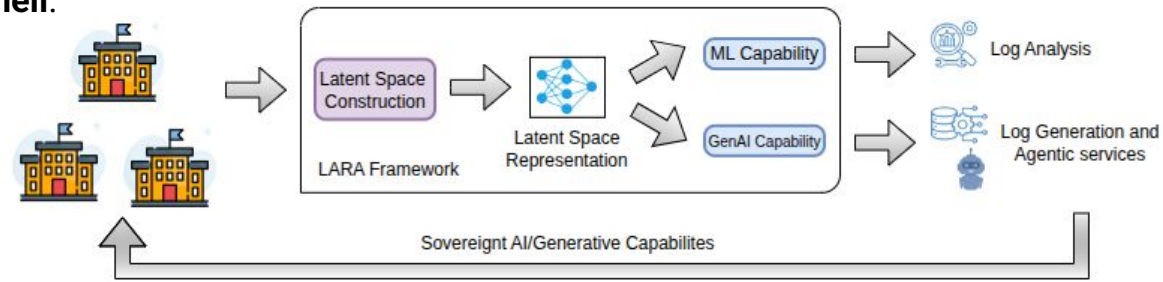
LARA project

Latent Spaces to empower NRENs with generative AI capabilities that prioritize digital sovereignty and resource efficiency.

Research Hypothesis:

- Small Foundational Models/Language models are cost-effective for focus-oriented tasks (fine-tuning).
- NRENS can benefit from this Small Foundational Models/Language models for different downstream tasks.
- Data Spaces: to envision how Latent spaces could be shared using Data Spaces.

The **LARA in a nutshell**:



Objectives

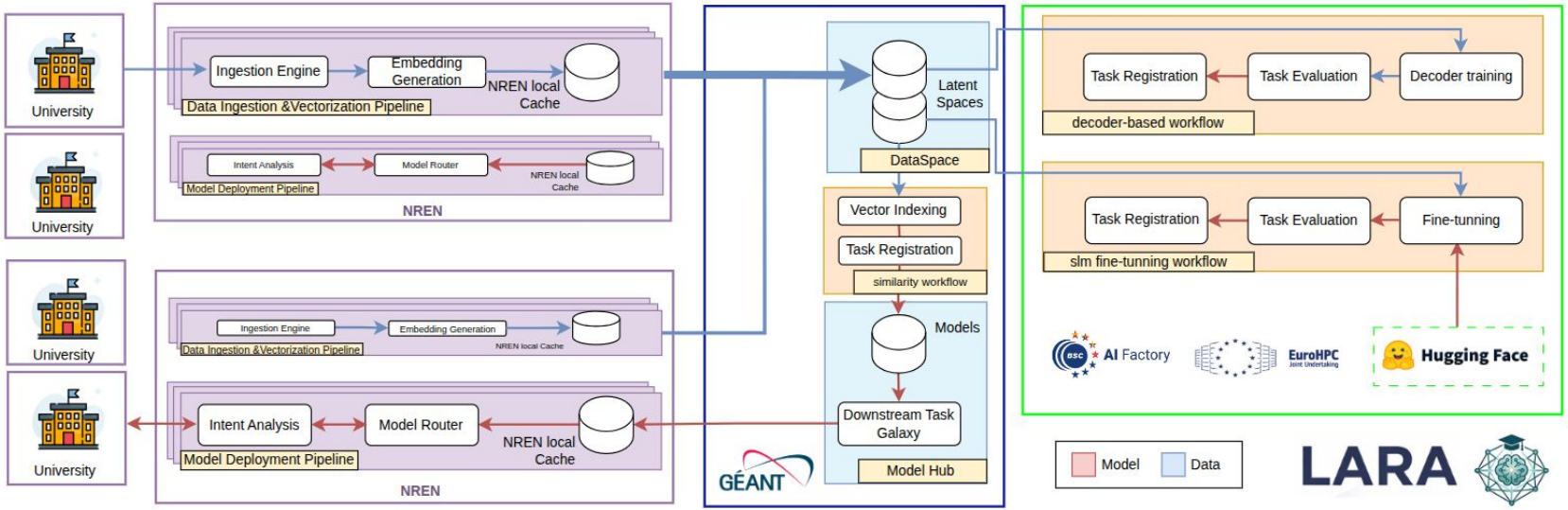
Objective 1 - Develop a Resource-Efficient Framework for Latent-Based Management
Design and implement a framework (LARA) capable of generating deep latent representations from multimodal application logs within NREN infrastructures.

Objective 2 - Scenario Validation
Validate the framework across two distinct real-world scenarios to assess its effectiveness, accuracy, and efficiency in network management and security monitoring tasks.

Objective 3 - Dissemination and Community Alignment
Actively engage the NREN community through publications and collaborative activities to ensure the framework aligns with community needs and is positioned for adoption.

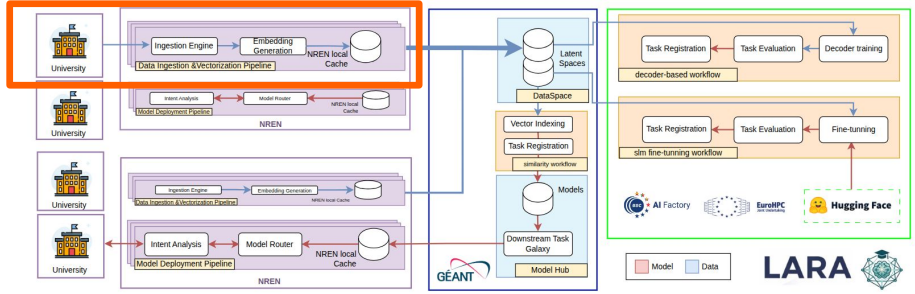
LARA architecture

From data to sovereign generative AI deployment



Albert Calvo, albert.calvo@i2cat.net **LinkedIn**/in/albertcalvo/

Methodology



Building the Latent Spaces (LS):

- From Raw data we built a compressed representation of the data, each datapoint is a vector.

  raw data (x) → encoder(x) → latent space z

- The encoder() is constructed using SLMs like BERT, Roberta for logs and text

- The different latent spaces are aggregated in an embedding store:

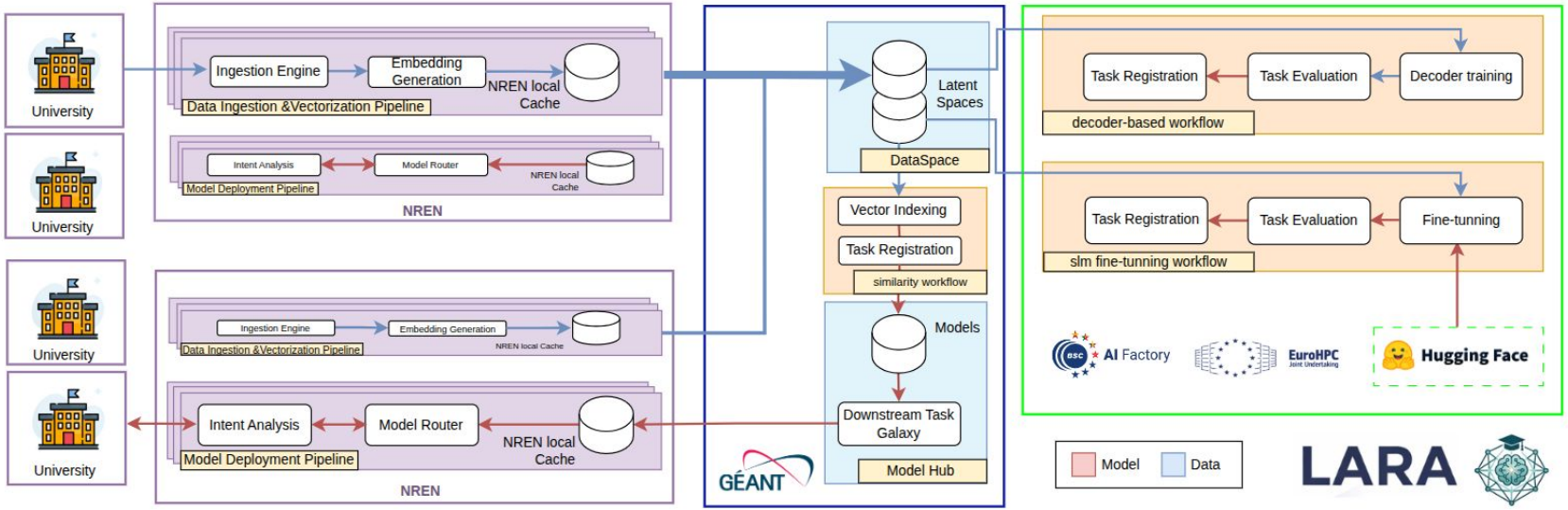
  The adoption of a dataspace enables privacy-preserving and secure data sharing across different NREN

  Bandwidth reduction (Latent space <<< raw data) .

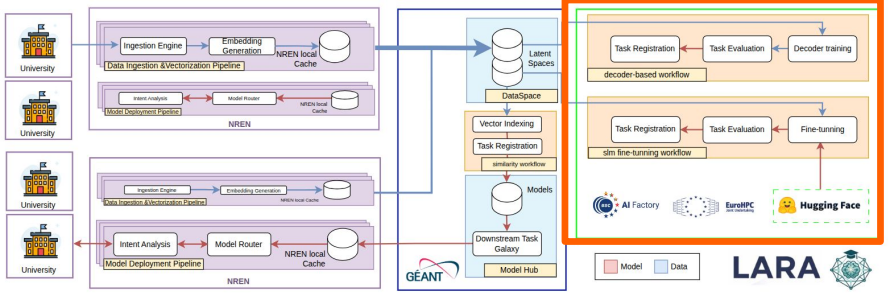
* All the gradients share the same consistent coordinate system

LARA architecture

From data to sovereign generative AI deployment



Methodology



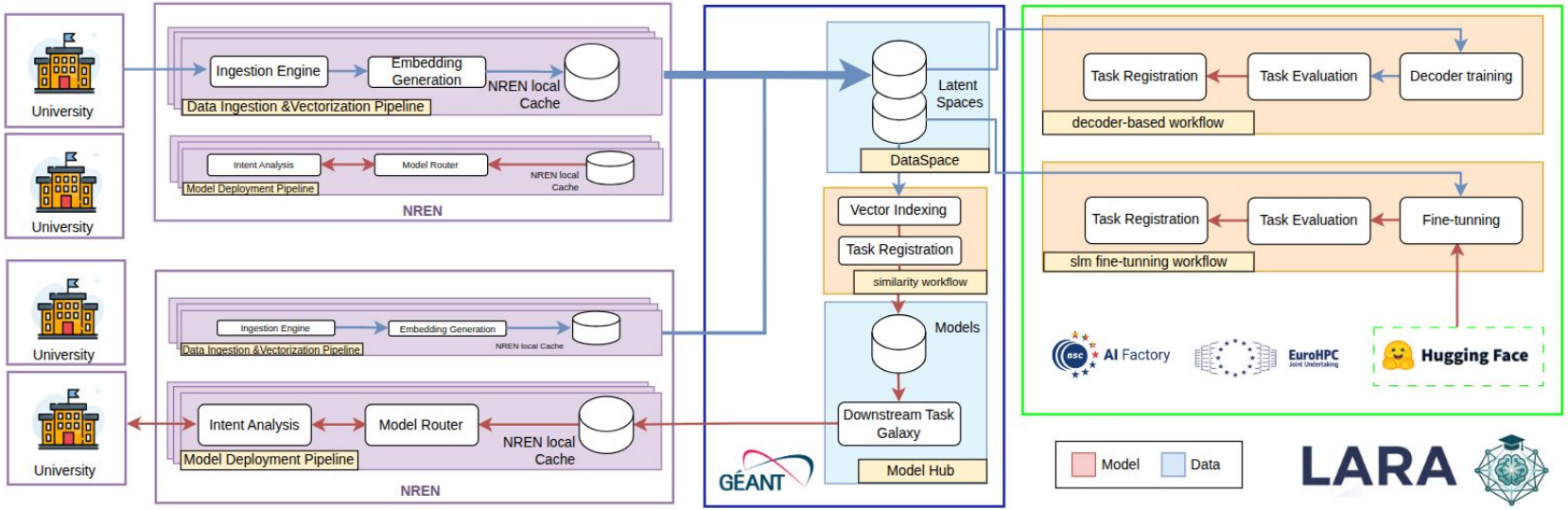
The previous LS are used for populate three workflows:

1. Similarity workflow -> the encoder produces embeddings from the data and a lightweight model uses those latent vectors to quickly predict labels or scores
   - *LS -> ML model -> Labels/Score*
2. Decoder-based workflow -> the encoder maps data into the latent space and a decoder model learns to reconstruct or generate data from those embeddings (e.g., synthetic logs or simulations).
   - *LS -> Decoder -> Reconstructed / Generated Data*
3. SLM fine-tuning workflow -> embeddings and retrieved context from the latent space are used to fine-tune a Small Language Model
   - *LS -> fine-tune a Small Language Model -> Generated Data*

✔ Applied for a grant to gain access to an AI factory

LARA architecture

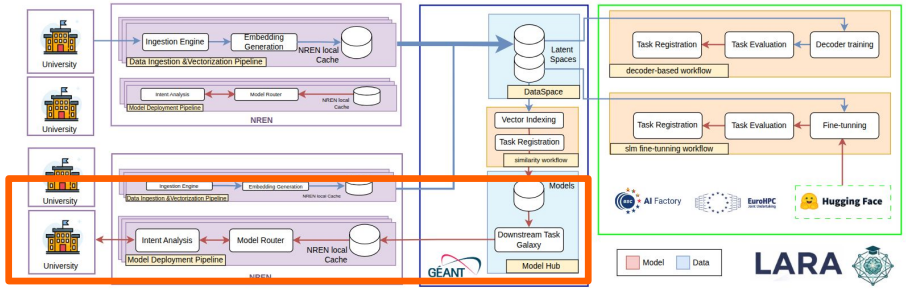
From data to sovereign generative AI deployment



Albert Calvo, albert.calvo@i2cat.net **LinkedIn**/in/albertcalvo/

Methodology

Finally, the NREN locally deploy a set of pre-trained tasks



Downstream Task Galaxy

Representation tasks
Tasks that discover the structure inside the embedding space

- **Similarity Search** — search within the latent space → event groping, network failures
- **Anomaly Detection** — detect outliers within the latent space → system monitoring, security alerts

Predictive tasks
Tasks that predict labels from embeddings.

- **Sequence Classification** — predict labels from embeddings → error type, triage
- **Sequence Prediction** — predict future events → incident prediction

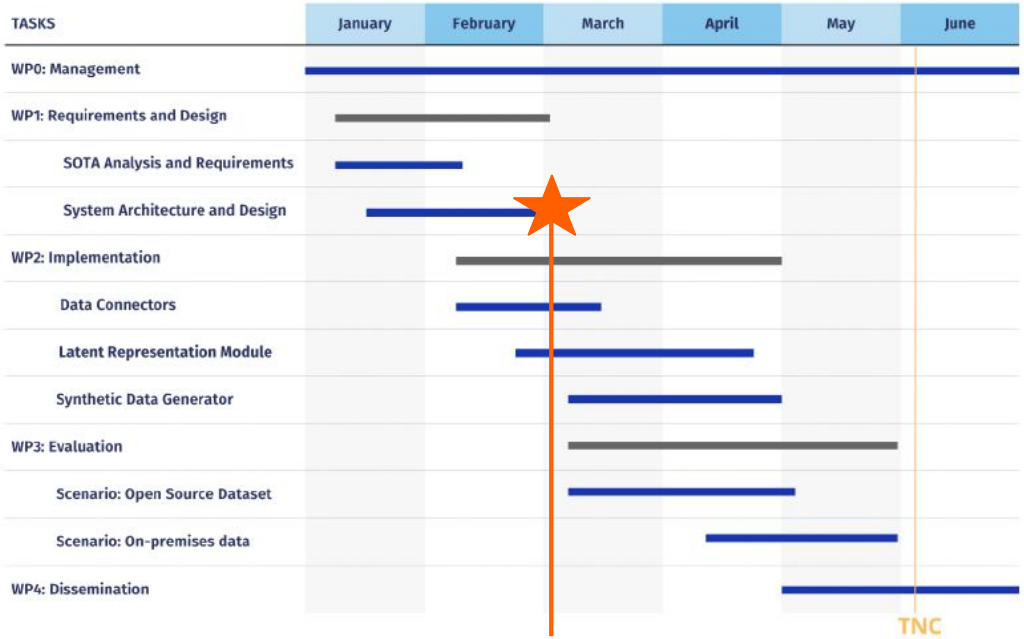
Generative tasks
Tasks that use embeddings to generate new content

- **Conditional Tasks** — use embeddings to generate new content → log generation (simulation), data augmentation
- **Reasoning/Agentic Tasks** — use embeddings for reasoning models → incident reports, summaries

i2cat

Albert Calvo, albert.calvo@i2cat.net **LinkedIn**/in/albertcalvo/

Timeline

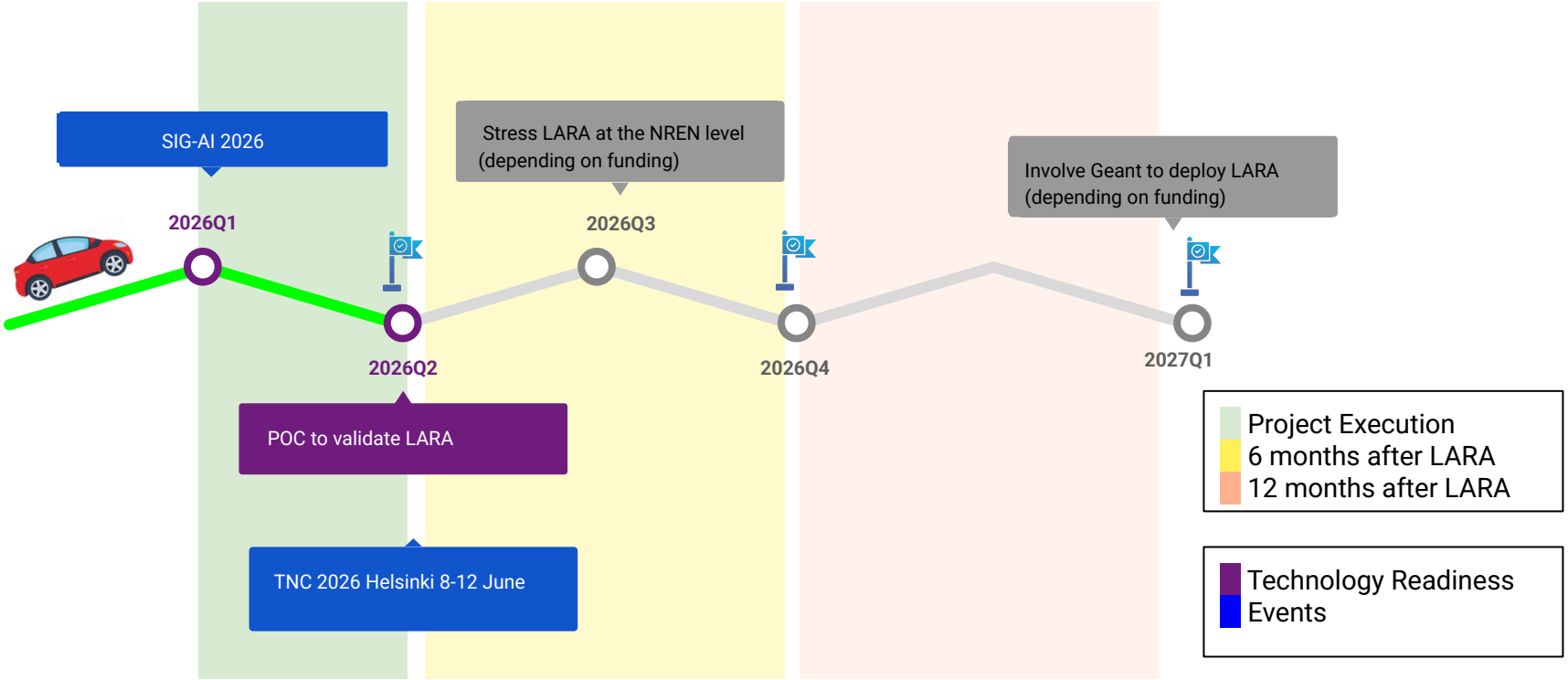


- WP0 Management (**On-going**)
- WP1 Requirements and Design (**Finished**)
- WP2 Implementation (**On-going**)
- WP3 Evaluation
- WP4 Dissemination

i2cat

Albert Calvo, albert.calvo@i2cat.net **LinkedIn**/in/albertcalvo/

Beyond LARA



Albert Calvo, albert.calvo@i2cat.net **LinkedIn**/in/albertcalvo/

Call to action

The opportunity for Sovereign generative AI is here!

- Explore the LARA technology
  - To explore together technological aspects of LARA
- Run a PoC
  - Deploy a small-scale implementation and see it working on your data and infrastructure.
- Join the ecosystem
  - Collaborate, experiment, and help shape sovereign AI in practice.

The LARA Ecosystem

i2cat

YOUR LOGO HERE

Scan me

i2cat **Albert Calvo,** albert.calvo@i2cat.net **LinkedIn**/in/albertcalvo/

i2cat

The Internet Research Centre

Thank you and see you in Helsinki!

Gran Capità, 2-4
Nexus I Building, 2nd Floor
08034 Barcelona
Tel. (+34) 935 532 510

X/Twitter | Linkedin | YouTube | Bluesky

Albert Calvo
albert.calvo@i2cat.net