# P16

## Communs numériques pour une IA souveraine

Inria

FRANCE 2030

RÉPUBLIQUE FRANÇAISE
*Liberté*
*Égalité*
*Fraternité*

# Notre mission

« Soutenir la mise en place et le développement d'une plateforme **souveraine**, **ouverte** et **interopérable** de librairies logicielles d'IA pour les entreprises françaises et son passage à l'échelle européenne » - *Mesure n°16 de la Stratégie nationale pour l'IA*

Inria

RÉPUBLIQUE FRANÇAISE
*Liberté*
*Égalité*
*Fraternité*

# Notre mission

P16
Communs numériques
pour une IA souveraine

## Développer

Développer ou soutenir le développement de bibliothèques logicielles

## Intégrer

Assurer une interopérabilité sans couture entre les bibliothèques

## Pérenniser

Maintenir ces communs à l'état de l'art scientifique et de la pratique technique

## Disséminer

Favoriser l’adoption dans le monde académique et industriel

# Notre référence, scikit-learn



scikit-learn est l'une des bibliothèques open source les plus populaires pour le machine learning, largement utilisée dans le monde académique et industriel.

**Impact**

- ✓ Adoptée par des milliers de chercheurs et de data scientists
- ✓ Pilier des workflows ML académiques et industriels
- ✓ Contributions massives et améliorations continues

- ⭐ **Etoiles:** 61.4k
- 🍴 **Forks:** 25.7k
- 👥 **Contributeurs :** 2,900+
- 🔄 **Commits:** 49,000+
- 🚀 **Releases:** 40+
- 📁 **Projets dépendants:** 1 million+
- 📦 **Packages dépendants :** 20,000+

## 🛠 Features

- ❑ **Classification:** SVM, Random Forest, k-NN...
- ❑ **Régression:** Linear Regression, Ridge, Lasso...
- ❑ **Clustering:** k-Means, DBSCAN...
- ❑ **Réduction de dimensionnalité:** PCA, LDA...
- ❑ **Sélection de modèle:** GridSearchCV, cross-validation…
- ❑ **Data pre-processing:** normalisation, encodage, imputation des valeurs manquantes…

# Notre montage unique

P16
Communs numériques
pour une IA souveraine

P16
Inria

- **Identification et validation** des librairies à intégrer
- **Pré-industrialisation** des librairies
- **Développement** de cas d'usage et applications métier
- **Collaborations** scientifiques et rayonnement de la recherche

:probabl.

- Entreprise à Mission de **Souveraineté Industrielle et Numérique**
- **Industrialisation** des librairies
- Offre commerciale
  - **Produits**
  - Professional **services**
  - **Formation** et **certification**

Inria

# Appel à manifestation d'intérêt

P16
Communs numériques
pour une IA souveraine

| Intégration dans le cycle de la donnée | Intérêt pour la souveraineté numérique | Résonance avec les besoins de l'écosystème |
|---|---|---|
| démontrer comment la bibliothèque contribue à la vision globale de P16 et peut s'intégrer ou étendre une des actions techniques (interopérabilité, préparation, apprentissage) | démontrer en quoi la bibliothèque contribue à renforcer la position nationale et européenne dans le domaine de l'IA (paysage concurrentiel, gouvernance …) | démontrer comment les solutions développées peuvent répondre à des cas d'usage spécifiques, ciblant des besoins industriels identifiés. |

**Généricité**

# Instruction des dossiers

P16
Communs numériques
pour une IA souveraine

Candidature d'une bibliothèque

Veille / état art
Benchmarks
Evaluation technique

**Instruction technique**

**Instruction stratégique**

Intérêt pour le domaine
Identification des besoins et use cases auprès des partenaires

Consultation des directeurs scientifiques

Recrutement d'ingénieurs
Pré-industrialisation

**A living document**

MAJ des benchmarks, alignement des use cases …

# Bibliothèques logicielles

P16
Communs numériques pour une IA souveraine

corese

## Intéropérabilité

Faciliter l'accès et le partage de données
Web sémantique & Knowledge graphs

skrub

## Data Preparation

Simplifier et Automatiser les tâches de préparation des données

scikit learn

:probabl.

## Machine Learning

Boîte à outils de Data Science

**Séries temporelles**

tslearn aeon

**Quantification d'incertitude**

MAPIE

**Explicabilité**

shapash

**Autres sujets d'intérêt**

Hybridation LLM / ML-DL, Apprentissage fédéré, Optimisation, Validation formelle, Reservoir computing, etc.

**Partenaires**

UNIVERSITÉ RENNES 2 — UNIVERSITÉ HAUTE-ALSACE — Capgemini invent — MAIF

Inria

# Bibliothèques logicielles

P16
Communs numériques pour une IA souveraine

## Intéropérabilité

Faciliter l'accès et le partage de données

Web sémantique & Knowledge graphs

## Data Preparation

Simplifier et Automatiser les tâches de préparation des données

## Machine Learning

Boîte à outils de Data Science

### Séries temporelles

tslearn aeon

### Quantification d'incertitude

MAPIE

### Explicabilité

shapash

### Autres sujets d'intérêt

Hybridation LLM / ML-DL, Apprentissage fédéré, Optimisation, Validation formelle, Reservoir computing, etc.

Partenaires

UNIVERSITÉ RENNES 2 — UNIVERSITÉ HAUTE-ALSACE — Capgemini invent — MAIF

# Skrub

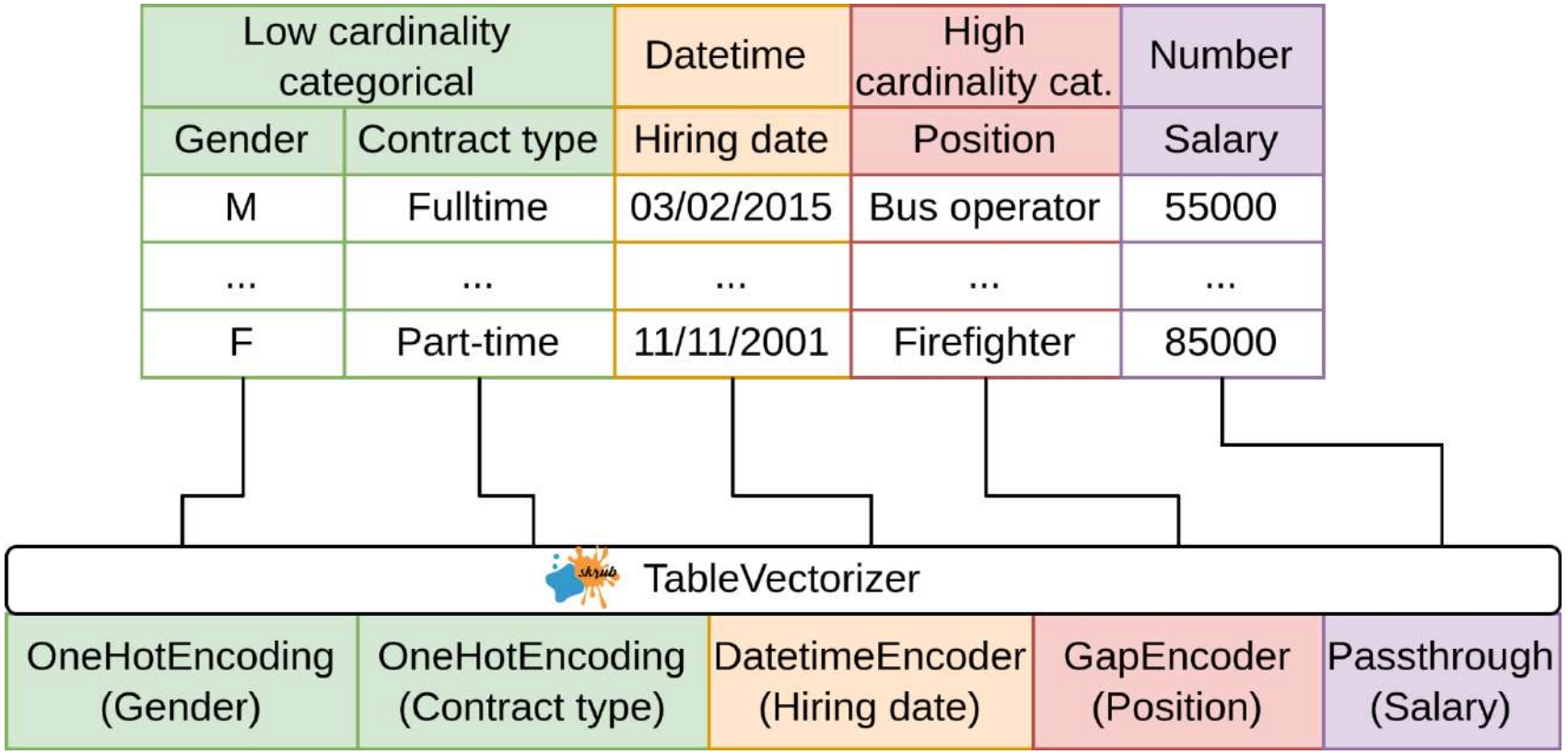
**Une bibliothèque Python dédiée au prétraitement des données tabulaires pour le Machine Learning**

- **Objectif** : Simplifier le nettoyage, l'enrichissement et la préparation des tables de données pour les modèles de ML.
- **Compatibilité** : Fonctionne avec les DataFrames de Pandas et Polars, et s'intègre parfaitement avec Scikit-learn.
- Open-source


**Cas d'usage**

- ❑ **Jointure entre tables :** relier des bases de données avec des correspondances imparfaites (ex. « Air France » et « Air-France ») → skrub.fuzzy_join
- ❑ **Analyse rapide de la qualité des données :** diagnostic automatique d'un DataFrame → skrub.table_report
- ❑ **Gestion des formats :** convertir automatiques les données en formats exploitables par les modèles → skrub.TableVectorizer

# Exemple 1 : TableVectorizer



# Exemple 2 : Augmenter les variables par jointure

| Movie title | Director | Cntr |
|---|---|---|
| E. T. | Steven Spielberg | USA |
| Your name | Makoto Shinkai | Japan |

| Title | Year | Rating |
|---|---|---|
| E. T. | 1982 | 7.9 |
| Your name | 2016 | 8.4 |

Joiner

| Country | Films |
|---|---|
| India | 1691 |
| United States | 1361 |
| Mexico | 1289 |
| Japan | 634 |

| Movie title | Director | Country | Year | Rating | Films |
|---|---|---|---|---|---|
| E. T. | Steven Spielberg | USA | 1982 | 7.9 | 1361 |
| Your name | Makoto Shinkai | Japan | 2016 | 8.4 | 634 |

# Exemple 2 : Augmenter les variables par jointure

| Title | Year | Rating |
|---|---|---|
| E. T. | 1982 | 7.9 |
| Your name | 2016 | 8.4 |

| Movie title | Director | Cntr |
|---|---|---|
| E. T. | Steven Spielberg | USA |
| Your name | Makoto Shinkai | Japan |

| Country | Films |
|---|---|
| India | 1691 |
| United States | 1361 |
| Mexico | 1289 |
| Japan | 634 |


Joiner

| Movie title | Director | Country | Year | Rating | Films |
|---|---|---|---|---|---|
| E. T. | Steven Spielberg | USA | 1982 | 7.9 | 1361 |
| Your name | Makoto Shinkai | Japan | 2016 | 8.4 | 634 |

```python
from skrub import Joiner

result = Joiner(aux_table, main_key="col1",
                aux_key="col2").fit_transform(main_table)
```

# Bibliothèques logicielles

P16
Communs numériques pour une IA souveraine

corese

## Intéropérabilité

Faciliter l'accès et le partage de données
Web sémantique & Knowledge graphs

skrub

## Data Preparation

Simplifier et Automatiser les tâches de préparation des données

scikit learn

:probabl.

## Machine Learning

Boîte à outils de Data Science

**Séries temporelles**

tslearn aeon

**Quantification d'incertitude**

MAPIE

**Explicabilité**

shapash

**Autres sujets d'intérêt**

Hybridation LLM / ML-DL, Apprentissage fédéré, Optimisation, Validation formelle, Reservoir computing, etc.

**Partenaires**

UNIVERSITÉ RENNES 2 — UNIVERSITÉ HAUTE-ALSACE — Capgemini invent — MAIF

Inria

# Bibliothèques logicielles

P16
Communs numériques pour une IA souveraine

corese

## Intéropérabilité

Faciliter l'accès et le partage de données

Web sémantique & Knowledge graphs

## Data Preparation

Simplifier et Automatiser les tâches de préparation des données

## Machine Learning

:probabl.

Boîte à outils de Data Science

**Séries temporelles**

tslearn

aeon

**Quantification d'incertitude**

MAPIE

**Explicabilité**

shapash

**Autres sujets d'intérêt**

Hybridation LLM / ML-DL, Apprentissage fédéré, Optimisation, Validation formelle, Reservoir computing, etc.

**Partenaires**

UNIVERSITÉ RENNES 2

UNIVERSITÉ HAUTE-ALSACE

Capgemini invent

MAIF

Inria

# Tslearn : ML pour séries temporelles

**Une librairie Python spécialisée**

- Outils optimisés pour l'analyse et le traitement des séries temporelles
- Compatible avec Scikit-learn
- Open-source


**Cas d'usage**

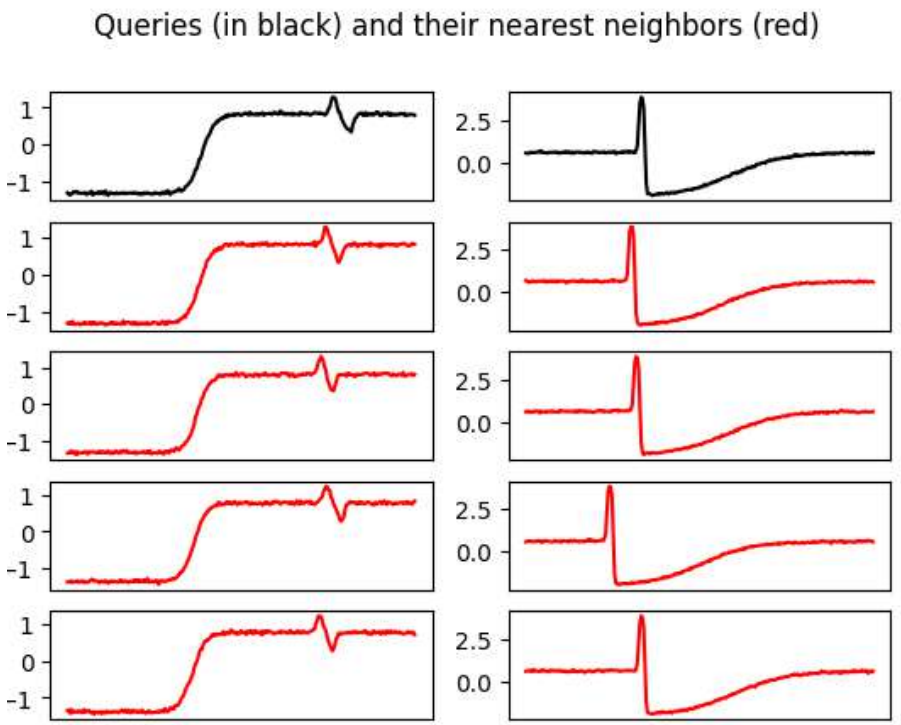
- ❑ **Clustering** : Regrouper des séries similaires (ex. clients avec comportements similaires) → tslearn.clustering.TimeSeriesKMeans
- ❑ **Classification** : Attribuer une étiquette prédéfinie à une série temporelle (ex. détection de pannes) → tslearn.neighbors.KNeighborsTimeSeriesClassifier
- ❑ **Alignement de séries** : Comparer des séries temporelles de longueurs différentes (ex. reconnaissance de gestes) → tslearn.metrics.dtw

# Exemple 1 : recherche de similarité



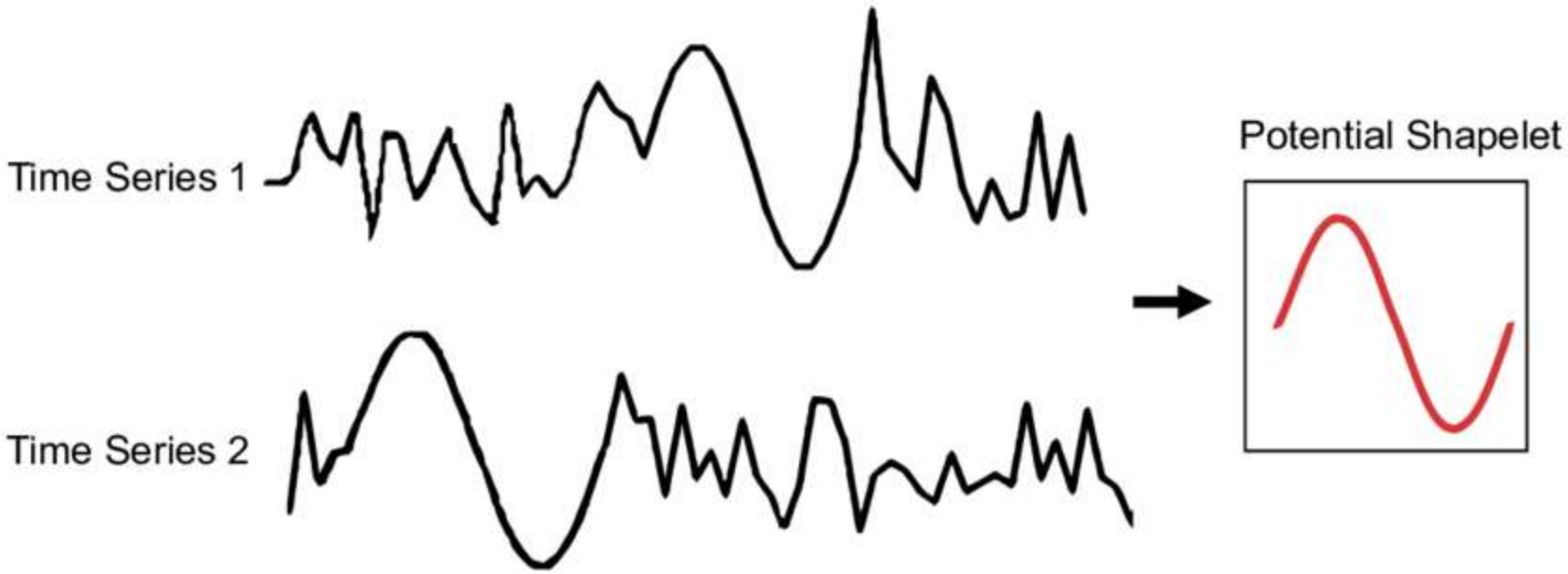
**Default distance : dtw**

```python
import numpy as np
from tslearn.neighbors import KNeighborsTimeSeriesRegressor
from tslearn.datasets import CachedDatasets

# Load the Trace dataset
X_train, y_train, X_test, y_test = CachedDatasets().load_dataset("Trace")

# Set up KNN regressor
n_neighbors = 4
knn = KNeighborsTimeSeriesRegressor(n_neighbors=n_neighbors)
knn.fit(X_train, y_train)

# Find nearest neighbors and make predictions for the first 2 test samples
n_queries = 2
predictions = knn.predict(X_test[:n_queries])
dist, ind = knn.kneighbors(X_test[:n_queries])
```



# What is a shapelet

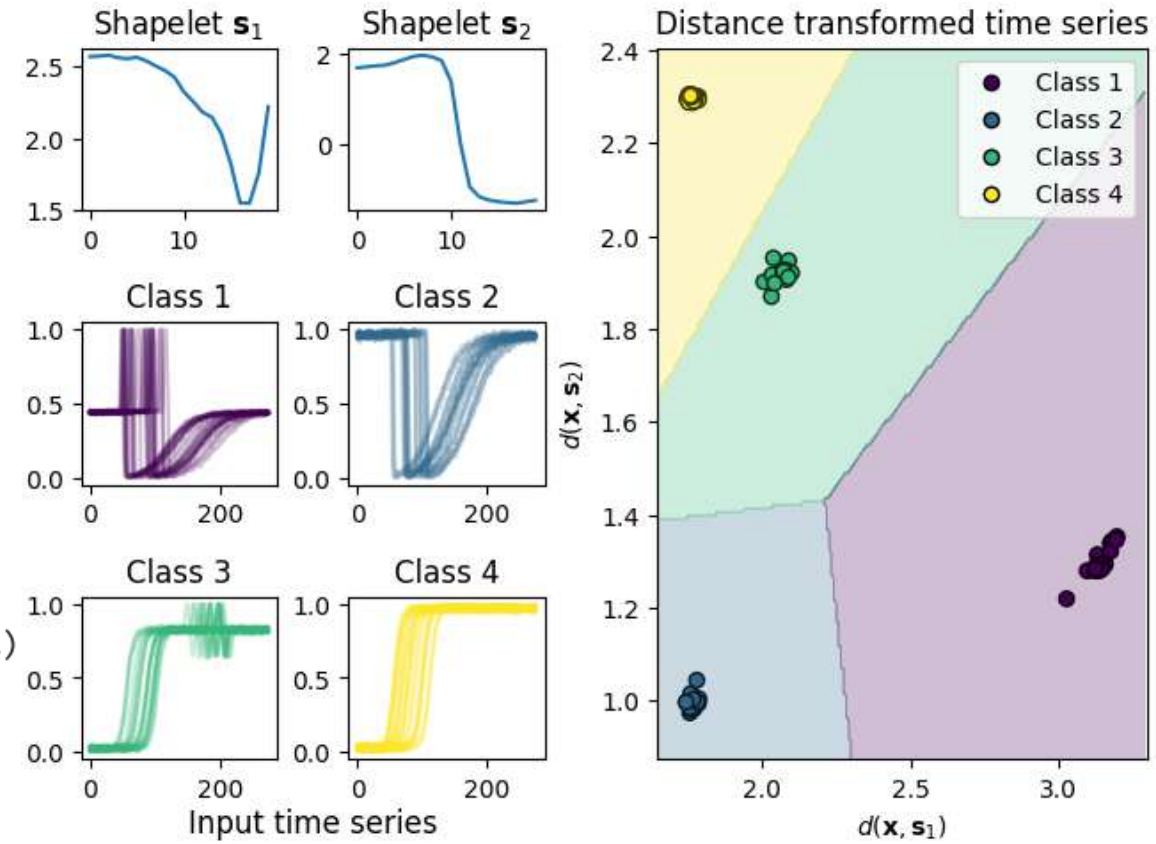




# Exemple 2 : tslearn, Keras & Shapelets …



```python
import numpy as np
from tslearn.datasets import CachedDatasets
from tslearn.preprocessing import TimeSeriesScalerMinMax
from tslearn.shapelets import LearningShapelets
from tensorflow.keras.optimizers import Adam

X_train, y_train, _, _ = CachedDatasets().load_dataset("Trace")
X_train = TimeSeriesScalerMinMax().fit_transform(X_train)
shapelet_sizes = {20: 2}  # 2 shapelets of length 20
model = LearningShapelets(n_shapelets_per_size=shapelet_sizes,
                          weight_regularizer=1e-4,
                          optimizer=Adam(0.01),
                          max_iter=300, scale=False, random_state=42)
model.fit(X_train, y_train)
```



# Autres exemples

P16
Communs numériques
pour une IA souveraine

- **Preprocessing and transformation** (normalization, dimensionality reduction, handling timeseries with variable lengths, subsequence extraction, etc.)
- **Summarizing the behavior of several series** using a robust average / barycenter (e.g., typical consumption profile)
- **Clustering** (e.g., grouping client profiles based on their energy-consumption)
- **Classification** (e.g., Electrical Appliance detection based on consumption patterns, anomaly detection, Grid failure causes )
- **Motif Extraction** (e.g. , identifying characteristic subsequences that allow for the distinction of classes)
- **Early classification** (e.g., early detection of a failure from the first signs in the series)

P16

Communs numériques
pour une IA souveraine

# Rejoignez notre Ecosystème

1 **Contact**

programme-ia.p16@inria.fr

2 **Website**

Visiter notre site :

p16.inria.fr