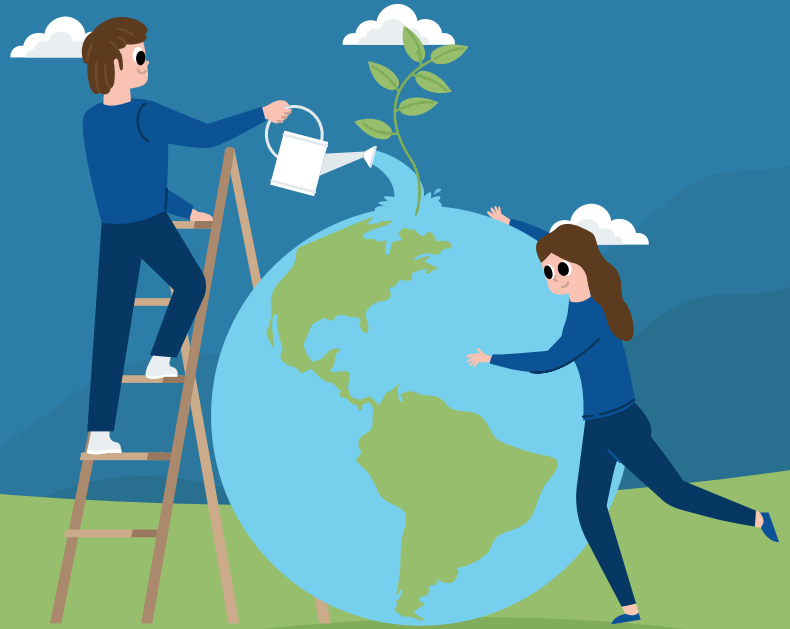
Max Cameron WU
Guillaume PHILIPPE

HACKATHON AI

In partnership with AFD and UNEP



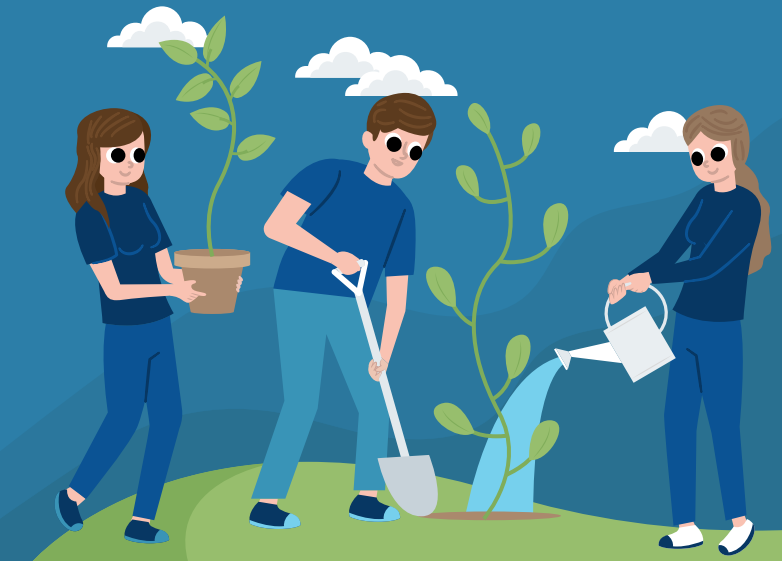
AFD
AGENCE FRANÇAISE DE DÉVELOPPEMENT

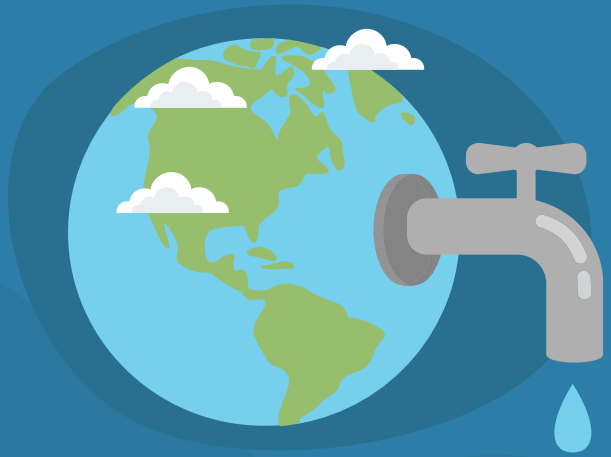
UNEP

isep
École d'ingénieurs du numérique

01 METHODOLOGY

How we decide to solve the challenge





IMPLEMENTATION CHOICE

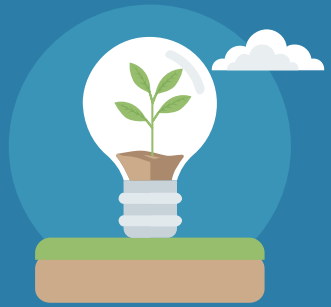
The system needs ...

... to be **data** free

... to be **scalable**

... to be easy to **deploy**

MODEL



FACEBOOK BART LARGE MNLI

Pre-trained NLI models as a ready-made zero-shot sequence classifiers

DISTILBART MNLI

Created using the No Teacher Distillation technique

DISTILBART MNLI

Premise is : 'Public transport is the most polluting form of transport because of the fleet constituted by old vehicles'

Hypothesis is : 'This text is about **public transport**'
Probability label is true : **98.82%**

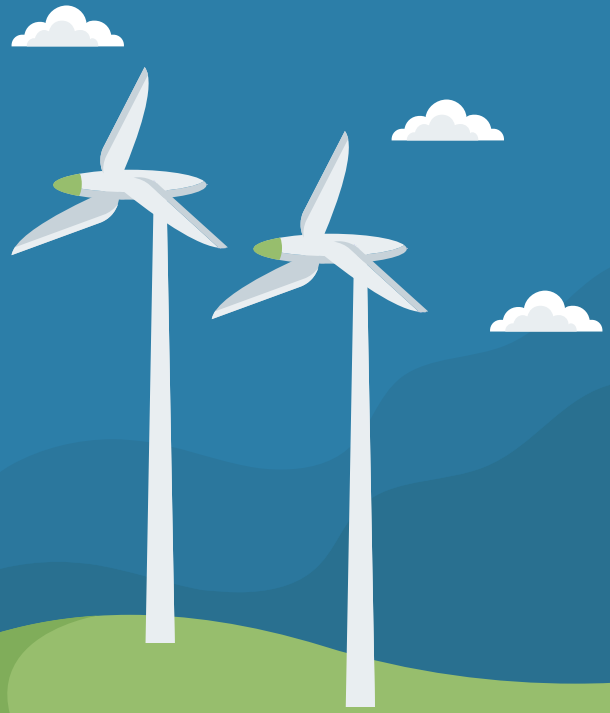
Hypothesis is : 'This text is about **climate change**'
Probability label is true : **74,59%**

Hypothesis is : 'This text is about **politics**'
Probability label is true : **1,32%**

- Choose the right hypothesis from the targets list
- Interpret properly the probabilities for each target

02
RESULTS

How our model performs



METRICS

F1 Score when target is among the X highest probabilities :

Top 1 : **0.19**

Top 3 : **0.45**

Top 5 : **0.57**

0 WATT

Consumption
for fine-tuning

0 SEC

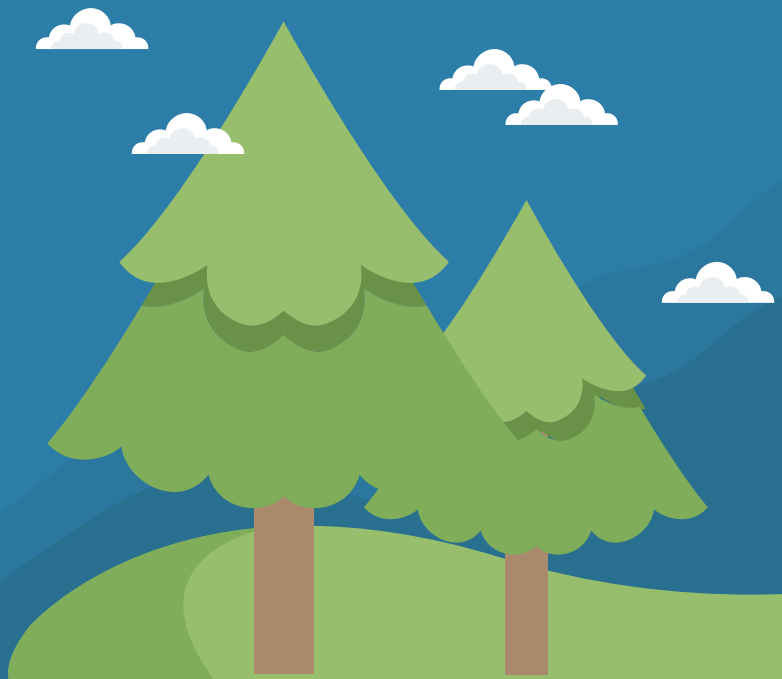
Fine-tuning
time

2 SEC

Inference time on CPU
Intel i5 2,3 GHz

03 DISCUSSIONS

What we can improve on our solution



IMPROVEMENTS

TARGETS

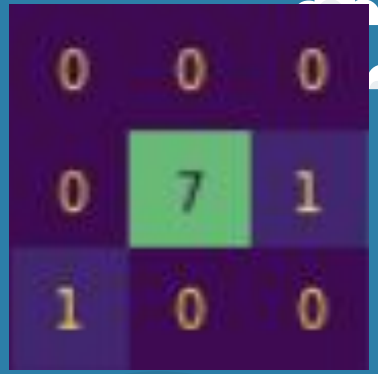
They must be more differentiable

COMBINE MODELS

The SDG meter can be used to obtain better results

CROSS SDG

e.g. in the dataset, target 15.1 is not assigned to a document of SGD 12



The matrix confusion shows the ambiguity between target 12.4 and 15.1 (both on pollution in nature)

THANKS!

https://github.com/gphilippee/hackathon-sustainable-development