SCOPING AND DESIGN REPORT: TECHNOLOGY TO SUPPORT DATA AVAILABILITY FOR RESPONSIBLE AI

Demonstrating practical use of Privacy Enhancing and adjacent technologies for well-governed data access for AI

Scoping and Design Report (Stage 1)

TABLE OF CONTENTS

1. Introduction

The Global Partnership on AI ("GPAI") is a multistakeholder initiative looking to educate and spread the benefits of AI amongst the public. It has been established with a mission to "support and guide the responsible adoption of AI that is grounded in human rights, inclusion, diversity, innovation, economic growth, and societal benefit while seeking to address the UN Sustainable Development Goals."

In the context of this overarching mission, Capgemini is supporting GPAI on a project aimed at disseminating processes and technologies that address concerns about privacy, sovereignty, IP protection, data security, data travel and localisation in AI applications. The project originates from the realisation that AI applications are limited by the lack of data availability caused by those concerns.

Technologies such as Privacy-Enhancing Technologies, or "PET" can be used to address the concerns, hence helping increase the availability of data needed by AI systems.

2. Objective

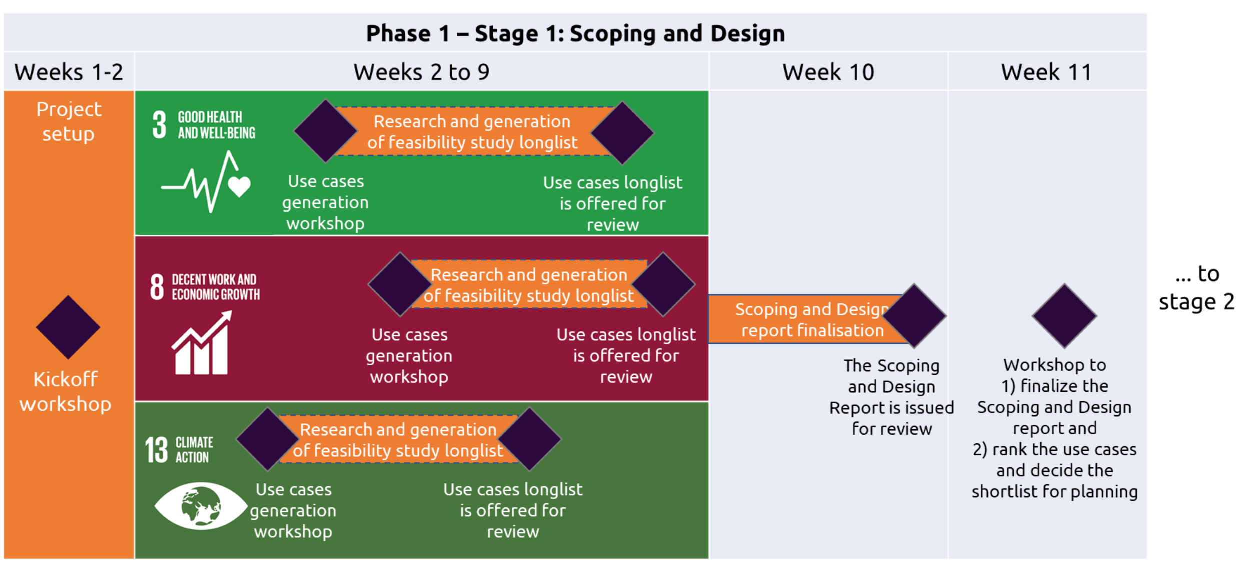
The project aims to identify one or more compelling use cases to make into a demonstrator project that can be used for dissemination to the public at a later stage. The demonstrator will showcase those processes and technologies applied to the public good, specifically in the context of three AI for good topics Climate Action, Future of Work and Better Health.

It is useful to note that the project aims to identify the best suitable use cases for *dissemination*, and not necessarily the ones offering the best *performance or impact* in addressing an issue. In other words, an extremely impactful AI application may not necessarily be the one that is the most suitable to educate the public and may not need sensitive data at all.

A successful report will enable the GPAI and any other relevant stakeholders to shortlist which use cases to bring further into planning, like the ones with the highest potential to demonstrate the processes and technologies and support outreach and develop awareness in the public.

3. Research Approach

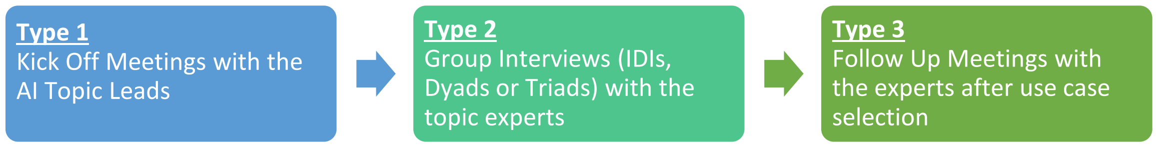
This report, and the workshop where it will be discussed, comp the first stage of the process, described in the picture below.



Because of the nature of the three AI for good topics of focus, the GPAI's topic leads, experts and other relevant stakeholders were interviewed in parallel to document opportunities across all three topics.

For each of the AI for good topics (better health, climate action and future of work), three different types of meetings/workshops were conducted. These meetings/workshops were held to gather more in-depth information from the GPAI's topic leads, subject matter experts, and other stakeholders.

Figure: Type of meetings/workshops



NOTE – *A-List* ***of Topic Leads and Experts*** *has been shared by GPAI for all the three AI topics.*

For the interviewees' reference, GPAI's definition of the elements of friction to data availability was offered during the meetings, as shown in the table below.

Table: Elements of Friction

Privacy	Identifiability or potential identifiability of individuals represented in data. While not typically considered in scope for 'privacy', there are similar concerns around the identifiability of organisations or of groups of individuals as well.
Sovereignty	Respecting company, government, community, individual and indigenous rights to govern and control use of data
IP Protection	Recognition of and respect for any property rights inherent in data
Data Security	Ensuring safe handling and housing of data throughout the data life cycle
Data Travel & Localisation	Legal requirements, size of data sets, and other considerations may limit the movement of data from their originating location

NOTE *-* ***Elements of Frictions*** *limiting the availability of data for AI applications have been defined by GPAI*

A series of criteria were defined to guide the evaluation of the use cases as they were collected from the interviews. The criteria belonged to three groups –

- Feasibility of the use case
- Relevance and Dissemination Potential
- Exploitability and Re-Usability

The first group was used as a "gating" criteria: if any of its conditions could not be verified, the use case was immediately discarded from further evaluation.

The criteria and how they were scored as described below.

Table: Use Case Criteria List

GPAI – Criteria		Score
1. Feasibility (Gating Criteria)	1.1. Is there a Candidate Owner for this particular use case? E.g., a GPAI ongoing initiative or partner?	**High:** If a suitable owner initiative exists within the GPAI portfolio of working groups and partners, and funding exists to support it, or there are plans for funding **Medium:** If there are no existing or future GPAI-funded initiatives, but GPAI members are aware of suitable similar initiatives funded by other entities such as government, academia, private sector etc. that could take ownership of the demonstrator project **Low:** Other
	1.2. Could the demonstrator project be Implemented in 15-18 months and within a budget of USD 1m total?	**High:** Yes **Medium:** Yes, with strong caveats (e.g., limiting the scope of the original use case) **Low:** No
	1.3. Is Data Available to enable the application of AI, as required by the use case?	**High:** Yes, sufficient data is available **Medium:** Some data is available **Low:** Very limited (or no) data is available
	1.4 What is the security risk, and how sensitive is the data? (Security risk: e.g., high if multiple parties involved, large "attack surface" Sensitivity: e.g., patient data (more sensitive) vs commute journeys (less)	**High:** Low sensitive data and low risk **Medium:** Medium sensitive data with manageable risk **Low:** Very sensitive data with high risk
2. Relevance and dissemination potential	2.1. Does this use case address one or more elements of concern limiting access to data, among privacy, IP protection, sovereignty, data security, data travel and localization?	**High:** 4 to 5 Elements **Medium:** 2 to 3 Elements **Low:** 1 Element (or Less à Disqualify)
	2.2. Does this use case have a sustainable positive societal impact? Will it stay relevant for about 1 - 2 years from now?	**High:** Yes, it will stay relevant longer & have a significant societal impact **Medium:** Yes, it will stay somewhat relevant for 1-2 years and have a medium societal impact **Low:** No, it will have little societal impact and won't stay relevant for that long
	2.3. Is this use case relevant across a range of geographies?	**High:** Yes, it is used across all countries **Medium:** Yes, it is usable in a few countries **Low:** No, it is useable within one or two countries only
	2.4 Does the use case demonstrate processes and technologies at scale with real data (subject to minimum security and privacy expectations in Q1.4)	**High:** Yes, the use case will demonstrate at scale with real-life data **Medium:** Yes, the use case will demonstrate processes and technologies but with limited/small-scale real-life data **Low:** Other

3. Exploitability and re-use	3.1. Building the demonstrator project will produce re-usable and generalizable knowledge (documentation, research...) and assets (e.g., datasets, source code...)	**High:** The project will generate re-usable and generalizable assets **Medium:** The project will generate re-usable assets **Low:** Other
	3.2. The use case shows awareness of dependence/ implications on data trust and data justice	**High:** the use case offers itself also to disseminate the topics of data trust and justice **Medium:** the use case offers itself also to disseminate either topic **Low:** no

The original methodology was intended to identify use cases only from GPAI-owned current and future research projects. The early interviews showed how this could not generate a significant number of options, mostly because many of the use cases did not have a clear owner, hence failing the gating criteria.

To address that, the GPAI Steering Committee asked the researchers to extend the scope of the investigation. Informally, the terms “Plan A” and “Plan B” were then used to refer to the fact that we were complementing the original approach (A) with additional options, as described in the following:

- **Plan A** – Any use case/project based on current and future work by GPAI
- **Plan B** - Any use case/project referred by experts that do not belong to GPAI
  OR Any company or organisation that has been working on similar kinds of technology and can be interested in collaborating with GPAI.

During the weekly status call, the GPAI Steering Committee could provide immediate feedback on the researchers' funding and guidance to focus effort on the use cases that were the most worth exploring further. In the later weeks of this stage, the use cases were re-categorised according to the three final categories below:

- **Category 1** - Use cases that have existing owners AND can address one or more elements of concern limiting access to data
- **Category 2** - Use cases that have potential owners AND can address one or more elements of concern limiting access to data
- **Category 3** - Use cases that do not have potential owners at this moment OR cannot address one or more elements of concern limiting access to data.

Category 3, by definition, could not pass the gating criteria and was not developed further.

4. Results

4.1. Overview

The use cases below have been detailed after a mutual agreement between GPAI and Capgemini:

Category 1

- Better Health - “Modelling the effects of human movement during a pandemic”

Category 2

- Climate Action – “Hyper-personal journey planning”
- Climate Action – “Hyper-personal city planning”
- Future of Work – “Smart cameras for ethical surveillance/monitoring”

4.2. General learnings and observations

The challenge of ownership

More than GPAI had expected at the beginning of the project, the challenge of finding a realistic owner for the demonstrator project demonstrated itself to be a common showstopper. It was easy to identify "categories" of owners (e.g., *a* municipality, *a* hospital…) but not a specific instance thereof.

The researchers believe that the cause for this may be in the nature of the GPAI working groups, which may develop their research extensively without the need to "channel" the findings and recommendations down to the context and needs of one specific user or organisation (*one specific* municipality, *one specific* hospital…) The involvement of these organisations may happen and is welcomed, but we understand that it is not systematic.

Suggestions/considerations for improvement: extend future working groups by inviting potential user organisations as "champions" for the category they belong to. By contributing so closely to the work, they will be comfortable with the research and finding and be the most likely owners of future implementations. Also, GPAI will have the additional benefit of the contribution that can validate the findings in the "real world".

Funding: opportunity or curse?

The problem of data availability for AI applications is, most commonly, framed as a "data giver" problem: resistance to sharing by the people or organisations that are described in the data. The way itself GPAIs define the element of friction – privacy, sovereignty etc. – puts the data givers at the centre of the argument.

However, anecdotal evidence from the execution of the project this far, shows as much friction by the "data takers", too: the institutions, researchers, and civil servants who would become custodians of the sensitive data. The concerns, in this case, are diametrical to the data givers', and are equally easy to understand lack of data literacy and skills that are suitable to take the responsibility for the data, reputational risk for the organisations, legal repercussions, fines, or any human error and cybersecurity incident that would break the delicate trust of the "data givers".

Particularly because the re-use of data is not exclusively a technology problem, the participants in a data-sharing arrangement need to feel the comfort of being supported in the non-technical aspects as well.

Suggestions/considerations for improvement: in case the process of this project had to be repeated, or, potentially, also for the successor project to this, we advise to offer the candidate owners not just the value of benefitting from the "gift" of the demonstrator (as a "technology asset") but also of support addressing directly the elements of concerns their side. E.g., legal advisory can make the data takers more comfortable taking the inevitable risk. Project management assistance can understand, describe, and manage that risk and communicate it to the stakeholders. Training may prepare and upskill teams as necessary to master what is needed to own the data-sharing initiative, etc.

The actual offer of services – rather than of *budget to cater for* services – is expected to be more welcomed, as it would avoid the task and complexity of procuring the service and controlling its budget.

5. The Use Cases

5.1. Better Health - "Modelling the effects of human movement during a pandemic"

5.1.1. Summary

Project Objective	Modelling the effects of human movement during a pandemic
Challenges Being Addressed	• Potentially disruptive effects on urban and transportation planning - People's mobility and way of life have drastically changed throughout the recent pandemic (COVID-19). • Planning non-pharmaceutical interventions (NPIs): any type of health intervention which is not primarily based on medication but includes actions that municipalities, individuals and households can take, to optimize the mobility during the pandemic and reduce its spread at the same time • Unforeseeable pandemic spread within hospitals and unexpected need for ward beds and ICU beds requirements • Alarming impact on the clinic's/hospital's ability to provide care due to the rapid transmission of the disease • Lacking knowledge about the healthcare requirements and correct treatment of the patient
Approach	Use AI to model a country's healthcare & mobility data and share the calibrated model with another country
Benefits in the real world.	Authorities can make optimum movement plans and restrictions to avoid pandemic spread and maintain normal living

5.1.2. Criteria Evaluation

GPAI – Criteria		Score	Remarks GPAI ***Expert Comments*** ***Capgemini's Analysis***
1. Feasibility (Gating Criteria)	1.1. Use Case **Owner**	High	GPAI • GPAI approved a budget of 120,000 Euros, which will cover activities until the beginning of 2023 • A joint project between Christian Mission for Development (CMD Africa), Cognizant, Oxford university & some Modelling Teams from New Zealand
	1.2. **Project Implementation**: Within 15-18 months and a budget of USD 1 million	Medium	• *The trial phase of training the AI model for pandemic data, of the first country, can be finished within the given time frame.* • *Collecting and introducing the mobility data to the AI model will take more than 15-18 months* • *Training & calibrating the model for other countries will take more than 18 months*

	1.3. **Data availability** to enable the application of AI	Medium	GPAI • Significant volume of sensitive medical data is available, but the private information is masked before it is shared with the AI models for training. • NZ govt is very particular in terms of data governance. They will not be willing to share any data till the time they get to see the benefit of sharing the data. This is primarily because of some indigenous people and local tribes who are not keen to share data even with the NZ government because of their colonialist roots • *Potential - Considering the current pandemic there is a high propensity that we can convince data providers to share the data for feeding in the AI application.* • *Mobility data is yet to be acquired for this Use Case*
	1.4 **Security risk and data sensitivity**	Low	GPAI Security Risk is High as the medical data includes very confidential and private information of the patients. This use case includes very sensitive data with a lot of private information • *Potential – NZ government might be interested in an AI model like this if technology can solve their data privacy concerns* • *Mobility data can be taken from:-* o *Transactional data – Very Sensitive* o *Telco Data – Medium Sensitive*
2. Relevance and dissemination potential	2.1. **Addresses elements of concern** that limit access to data	Medium to High	• *Patients' anonymized private data will indeed be shared by private and public healthcare departments with a central AI model governing body, raising concerns about "Data Travel and Localization" and "Data Security."* • *Mobility data also has similar elements of concern*
	2.2. Have a **sustainable positive societal impact** (for 1-2 years from now)	High	GPAI In future, this model can be used for unforeseen pandemics situations (which are similar), and also for combatting general healthcare challenges as well
	2.3. **Relevant across geographies**	High	GPAI • The core concept of this use case is to create and train an AI model on a Country's healthcare & mobility data and share the calibrated model with another country.

			• The end objective is to cover all the countries throughout the world by transfer of information
	2.4 **Demonstrates processes and technologies** with real data	High	GPAI This use case is currently being investigated. Real data is being shared by private firms like Oxford and governmental entities like the Rwandan government, for testing purposes.
3. Exploitability and re-use	3.1. Produce **Re-usable** and generalizable knowledge	High	GPAI • This use case is used to train the AI model to fight against pandemic healthcare challenges & general healthcare challenges as well. • PET addresses most of the elements of concerns (“Data Privacy”, "Data Travel and Localization" and “Data security.") mostly for mobility-related data. • Most of the output that this AI model will provide can be used across geographies considering that healthcare challenges rising due to the pandemic will be more or less the same
	3.2. Awareness of dependence/ implications on **data trust and data justice**	High	*The calibration of an AI model is primarily reliant on high-quality and large-volume data. To profit from this AI Model, each country must contribute high-quality data.*

5.1.3. Additional Information

5.1.3.1. AI System Details

Currently, the modelling approach is divided into two parts:

- Epidemic modelling
- Economic modelling

This approach calibrates AI models over different datasets (historical), a set of global parameters, and a set of local parameters. A multi-objective genetic algorithm is used in this AI model.

5.1.3.2. Partnership and Association

- Data coming from university like oxford is in open source, there is no licensing involved.
- The branching model has been paid by the New Zealand government and there are no plans to commercialise it.

5.1.3.3. Data Availability

Cognizant has been using the Oxford data set which is in open. It's a standardised dataset from several countries, and it's also heavy on deep learning. They're looking at the neural network to determine what the parameters should be, and then trying to get it to match to the again the spread curves.

5.1.3.4. Data Barriers

- The branching model in New Zealand uses data from New Zealand that is not publicly available. The government has a lot of information that will not be available outside NZ.
- Need to check the level of accuracy if we use the standardised oxford data sets in the branching model.

- This model has not been exposed to large datasets so far, if we take it to a new country and add data, we have to check the quality of output.

5.1.3.5. Type of Datasets

There is a base transmission rate which is governed by the crown of the virus which varies across different places. It can be determined by human movement and how often they interact in those locations. For instance: it doesn't spread as readily in rural regions where the population is not very dense.

Mobility data capturing human movement will be important specifically in shared facilities like petrol stations, and supermarkets, particularly when the lockdowns are happening.

Cell phone data can't be used because of privacy concerns, and in certain situations, it can be government data that isn't truly publicly available.

5.1.3.6. Challenges accessing data

- Cell phone data cannot be accessed in NZ because of privacy concerns.
- Transportation data might be accessible as it does not have a high level of privacy concerns like where one lives, or where he/she works. To improve the efficiency of public transportation, data should be acquired.
- When one user swipes the card at the supermarket and another user swipes their card at the same moment, it can be confirmed they're both at the same place at the same time. As a result, the chances of people distributing COVID to each other are significantly higher compared to if they come across gas stations where no physical contact is possible.

5.1.3.7. AI Modelling Infrastructure

Cognizant staff mostly runs on the cloud, New Zealand modellers run on private infrastructure (like private VMS) due to data privacy concerns. In most cases, it will be a combination of both considering most of the countries have restrictions on health data not supposed to leave the country.

5.1.3.8. Future Concerns

- Need to check how can the models (that exist already) use this common data set from Oxford.
- Need to identify that if they are just restricted to that data set, how much do they lose in terms of the accuracy of prediction.
- By using standardized datasets across multiple countries, do we improve our models or by losing that sort of proprietary mobility data that's helping inform some of the models, do we lose accuracy.
- If we lose too much accuracy, then there will be a fundamental requirement for technology (like PET).

5.1.4. Conclusions

5.1.4.1. Challenges we observed in our selection and analysis

- Here the AI Model needs to be calibrated. The team is not sure whether the AI model has been calibrated or not. Oxford works in the open data source they may have tried to calibrate the model with their standardised datasets.
- Data sharing also seems to be a concern, no one is willing to share the data.
- Travelling of data outside the country is a constraint here. AI model is travelling here but data cannot travel. Once you receive the AI model, it needs to be calibrated with your data.
- Also, there is no immediate need for technology (like PET), as they are not using any mobility data like telco data, banking transaction data etc.

5.2. Climate Action - "Hyper-personal journey planning"

5.2.1. Summary

Project Objective	Minimise the environmental impact of private transport by promoting extremely personalised and customised – though eco-friendly – public transport, carpooling, and other forms of transport
Challenges Being Addressed	• Creation of shared transportation policies. • Growing adoption of bicycles and other ecologically friendly forms of transportation. • Numerous large-scale movements to build a better infrastructure for public transportation. • Promote environmentally-friendly transportation while reducing emissions. • To gather a lot of private travel information and safeguard it
Approach	• Experiment / build a next-generation journey planner tool that suggests passenger journeys that are deeply integrated with the needs of the individual while – at the same time – optimising them to achieve the challenges. • The elements of hyper-personal data that could be used are their real-time location, their agenda, their general needs (e.g. duties as a parent), any mobility limitations or impairments etc.
Benefits in the real world	Reducing CO2 emissions while improving the living standards and health of the citizens

5.2.2. Criteria Evaluation

GPAI – Criteria		**Score**	**Remarks** GPAI *Expert Comments* — *Capgemini's Analysis*
1. Feasibility (Gating Criteria)	1.1. Use Case **Owner**	High	GPAI • Toulouse's current projects (non-GPAI) are funded by the European Commission and also by the French Government • Any relevant municipality (inner-city travel) or metropolitan area (more typical of the daily commute in and out of cities) can be potential owners *The interviewees named specifically the French government and the municipalities of Paris or Toulouse. The French government plans to invest 40 million EUR to address climate change using AI.*
	1.2. **Project Implementation**: Within 15-18 months and a budget of USD 1 million	Medium	• *Yes, reducing the environmental impact of travel is compelling and – unfortunately - will not be solved sooner than the 2 years dissemination target.* • *The demonstrator project could be developed to handle a portion of the bigger use case, depending on the complexity of the viability and the scale of the area being addressed.*

	1.3. **Data availability** to enable the application of AI	Medium to High	**GPAI** Toulouse metropolis has a **partnership with several mobility data providers** *Telecom operators and other data providers can also provide traffic data that can be used to understand* *a) the travelling habits of individuals and* *b) general characteristics and patterns of the viability infrastructure of a geography*
	1.4 **Security risk and data sensitivity**	Medium	**GPAI** Individuals' movements, commute habits, and the location of their homes or offices are likely to be considered sensitive data, but not as highly sensitive as medical or financial records. *When people use Apple or Google's traditional journey planners on their smartphones, they already share this type of information regularly*.
2. Relevance and dissemination potential	2.1. **Addresses elements of concern** that limit access to data	Medium to High	**GPAI** Yes, data privacy, sovereignty, and data security are of concern.
	2.2. Have a **sustainable positive societal impact** (for 1-2 years from now)	High	• *AI provides suggestions and plans the daily tasks of an individual by analysing the city traffic, fellow travellers on the same route (for carpooling), and other activities related to the users' requirements.* • *Using AI for making the Car sharing/public transportation more convenient and personalized for use* • *Reducing the use of private vehicles and promoting sustainable modes of transport like cycling.* • *Trying to encourage people for choosing the emission-free transport mode.* • *Motivating commuters to take the shared mode of transport like cars, bikes etc.*
	2.3. **Relevant across geographies**	High	• *Yes, the problems dealt with by this use case are common to most areas of the world.* • *Besides, it is not just cities, but rural areas, too, particularly considering the general issue of the population getting older and requiring smarter ways to travel*

	2.4 **Demonstrates processes and technologies** with real data	High	• *Yes, traditional journey planning systems (Google, Apple) do already use real data at scale, though without adding the extra degree of personal data we envisage in this use case.* • *Only the need to fit within time and budget may suggest reducing the scope of the demonstrator in size.*
3. Exploitability and re-use	3.1. Produce **Re-usable** and generalizable knowledge	High	*Yes, as long as the demonstrator does not "overfit" the characteristics of the geography it is built for.* *E.g., we may argue that a journey planner optimised for Amsterdam or Venice, may not be effective for another major city*
	3.2. Awareness of dependence/ implications on **data trust and data justice**	High	*Public transport is usually the only option for the less wealthy parts of society. The implementation of the use case should not put parts of the population in a condition where they need to be forced to share their data to use public transport if they do not feel to. Using public transport should still be possible.*

5.2.3. Additional Information

5.2.3.1. Project Ownership

- Currently Toulouse is engaged in multiple projects relating to "Hyper-personal transportation and traffic management."
- One of the projects was funded by the European Commission, while the other one was funded by the French government, with all funds dedicated to the project's goal.
- If Toulouse and GPAI collaborate to develop an AI and technology demonstrator, it might not be included in the current projects, but rather be considered a standalone project with its funding.
- Toulouse is already working on a project, indicating that they are putting significant effort. It feels like a positive premise that many partners are already working together in a group.
- Toulouse Official believes that funding is required for such projects, and they could help raise funding from the Toulouse municipality and the French government.

5.2.3.2. Project Implementation

There are two sub-projects related to Hyper-personal journey planning being operated by Toulouse Municipality: -

- The first project was completed in 2021, and the transport authority is currently testing new technology created in the first project.
- The second project is currently being worked on and will end in December 2023.

5.2.3.3. Data Availability

Toulouse has a partnership with data-providing players such as Orange and TomTom (private data holders), Public Transportation Operators and Private Mobility Service Provider (a carpooling service providing company), as well as infrastructure information providers such as open street maps and so on.

Toulouse officials already have the following: -

- ➢ Have a proper agreement with data owners
- ➢ Well-structured data storage

5.2.3.4. Security risk and data sensitivity

Bosque Didier (Innovation Manager & Expert from Toulouse): –

- "At the beginning of a project, stakeholders are often not aware of the value add they can derive from the data-sharing platform. As a result, we provide technologies that enable the agile integration of data sources."
- "Currently there are tools which allow companies to anonymize the information for data sharing system, and clustering information among other data handling services."
- "We are GDPR compliant because we don't use Privacy information on the data sharing, but maybe I can say that in the projects that Toulouse & GPAI can define together could be interesting to go beyond the strategy by using technology and to identify if technology (like PET) is more efficient than an approach based on anonymization."

5.2.4. Conclusions

5.2.4.1. Challenges we have observed in our selection and analysis

- GPAI has not funded the project
- The models are owned by Toulouse Municipality & other private companies.
- Funding of the current ongoing is done by the European Commission and the French Government
- Currently organisations don't use private information in data sharing. Maybe in the future, the projects that Toulouse & GPAI can define together could be interesting to go beyond the current strategy by using technologies.

5.3. Climate Action - "Hyper-personal city planning"

This use case is an extension of the "Hyper-personal journey planning" use case.

The same data as in the "Hyper-personal journey planning" use case can be used to build additional AI models to support municipalities or local administrations in their planning activities. This could support the local authorities both in their short-term planning (deployment of fewer or more buses, temporary traffic alterations…) and medium-to-long term planning (areas to allocate to the building of shopping centres, schools, hospitals, or the revision of the effectiveness of the road network…).

5.3.1. Summary

Project Objective	Augment and improve the effectiveness of the planning of planning cities or wider geographical areas, thanks to an extremely granular awareness of citizens' and commuters' needs.
Challenges Being Addressed	• Work together with the municipal authorities to provide the best infrastructure or strategies for reducing traffic. • Creating an easily accessible infrastructure of a city, and eventually reducing transportation needs • Creating the ideal bicycle route and determining "how many people will be keen to take it" • Numerous large-scale movements to build a better infrastructure for public transportation. • Transportation planning while the population growth of a city is increasing. E.g., every ten years, the French census uses traditional methods of gathering data on people's mobility needs. On the other hand, the population of major cities like Toulouse is growing too quickly for that frequency to be useful (15,000 people/year on average).
Approach	Augment more traditional city planning techniques by using AI and extremely granular data describing the needs of citizens and commuters.
Benefits in the real world	Reducing CO2 emissions while improving the living standards and health of the citizens

5.3.2. Criteria Evaluation

Same as for "Hyper-personal city planning".

In terms of suitability for dissemination, the "Hyper-personal city planning" offers a higher degree of exploitability. The public can understand well the experience of travelling in an urban area or commuting, and will immediately envisage the benefits of sharing their personal data. City planning is, instead, a more sophisticated, complex activity that delivers its benefits more subtly and within longer time horizons.

5.4. Future of Work - "Smart cameras for ethical surveillance/monitoring"

5.4.1. Summary

Project Objective	Use AI-powered cameras for the ethical surveillance and monitoring of outdoor or work environments and intervene in the case specified events are detected (e.g., hazards) without the involvement of human operators
Challenges Being Addressed	• Monitoring an environment for accidents or security threats is often necessary. However, the use of human operators to monitor the video being produced offers itself to potential abuse: the video may be distributed and stored on unsecured networks and systems, or accessed by unauthorised personnel • Such systems may also possibly be the only solution to scale to cover large geographical areas without being intrusive on citizens or workers (during the interviews, for example, the example of Paris as the next Summer Olympics city was often referenced).
Approach	An AI running within the camera will raise an alarm when it detects relevant events, and the necessary actions will be made right away to resolve the situation to assure security.
Benefits in the real world	• Surveillance and monitoring of environments can be implemented/extended without violating the privacy of the people or confidentiality of the organisations being involved and, at the same time, maximising cost-effectiveness, and scalability. • Better safety for citizens/workers etc.

5.4.2. Criteria Evaluation

GPAI – Criteria		Score	Remarks GPAI *Expert Comments* / *Capgemini's Analysis*
1. Feasibility (Gating Criteria)	1.1. Use Case **Owner**	Medium to High	GPAI Start-Up called "22" is currently working on a project to create solutions for French government bodies.
	1.2. **Project Implementation**: Within 15-18 months and a budget of USD 1 million	Medium	• *Planning to use this technology for a wider audience expected before Paris Olympics 2024*
	1.3. **Data availability** to enable the application of AI	High	• *Data is going to be generated by the general population from Smart CCTV Cameras (real-time video footage)* • *Smart cameras can be installed anywhere in the world* • *This AI system can also use simulated/synthetic data for defining rules and model training*
	1.4 **Security risk and data sensitivity**	Medium to High	GPAI Using technology, humans are not identified (From Smart CCTV) but are considered as an object/indicator. No one can have access to the footage, but only the indicators.

			Sensitive data is not generated but technology will be needed for preventing the sensitive data to be produced, to prevent later the unethical use of the same
2. Relevance and dissemination potential	2.1. **Addresses elements of concern** that limit access to data	High	• *Generation of sensitive private data needs to be prevented* • *IP Protection is required* • *Sovereignty can be a major concern which needs proper security to be addressed* • *A lot of data will be travelling from smart cameras situated in different parts of the cities to a central location for administration*
	2.2. Have a **sustainable positive societal impact** (for 1-2 years from now)	Medium to High	• *This use case can improve law and order in the society* • *It can help to improve disaster/emergency management.*
	2.3. **Relevant across geographies**	High	• *This Use Case uses real-time data to be trained, therefore smart cameras can be installed anywhere in the world.* • *Situational learnings from a location can be calibrated into new geographies*
	2.4 **Demonstrates processes and technologies** with real data	High	• *This use case can only start by using real-time video footage provided by the Smart CCTV cameras.* • *Processes and technologies can use real data for trials and training.*
3. Exploitability and re-use	3.1. Produce **Re-usable** and generalizable knowledge	Medium to High	*PET which will get implemented in this use case can be reused for other AI applications concerning law and order and throughout several geographies*
	3.2. Awareness of dependence/ implications on **data trust and data justice**	High	*A government/monitoring body should provide good volume and quality of data to calibrate the AI Model according to its location and extract the maximum benefit from the AI*

5.4.3. Conclusions

5.4.3.1. Challenges we observed in our selection and analysis

None worth documenting.

6. Appendix –

The 25 use cases listed below are the result of the full set of discussions with the topic leads and experts from GPAI and other stakeholders, based on the research approach described earlier in this document.

Pandemic Modelling

The pandemic modelling use case in particular has developed through a series of transformations. The original use case (“1”) was detailed into three independent use cases (“1,1”, “1,2” and “1,3”) the last of which only met the gating criteria.

Finally, following GPAI’s directions, a new use case “1B” was created by merging “1,1” and “1,3” and adding an extra element of sensitive data (originally not required by 1,1) to augment the potential of the original 1,3 even further.

SL No	Topic	Use Case	Category
1	Better Health	**Pandemic Modelling -** Developing evidence-based models to support healthcare decisions and avoid negative economic implications while handling pandemic situations.	3
1.1		**Calibration of data describing pandemics –** To be done across different datasets in different geographies	3
1.2		**Pandemic Modelling: The Economic Impact of Pandemics -** Build a model for checking the economic impact during a pandemic	3
1.3		**Modelling the effects of human movement during a pandemic -** The spread of pandemics is closely linked to the movement of people and their gathering, typically in public places.	1
1B		**Modelling the effects of human movement during a pandemic –** The spread of pandemics is closely linked to the movement of people and their gathering, typically in public places. Detailed visibility of individuals’ movement – without the use of de-identification techniques – may further improve the effectiveness of the analysis.	1
2		**Surgical Schedule Optimization** - Using AI to automate the surgical procedures and support the surgeon to follow the schedule with SOP assistance	3
3		**AI for image – CT & Xray Scan** - AI prediction system to enable better evaluation of CT scan to report analysis	3
4		**AI to validate assumptions and replicate medical trial setting -** Building a network of like-minded real-jworld evidence side, which will be working on patient data (collected during patient care) to validate assumptions and replicate trial setting, without setting up a trial.	3
5	Climate Action	**Hyper-personal journey planning** - AI provides information about the city traffic, fellow travellers on the same route (for carpooling), task planning and other activities depending on your travel plans.	2
6		**Hyper-personal city planning** - Changing the traditional city structures and replacing them with a "smart city plan" providing all the required amenities (access to public transport, shopping mall, grocery shops/supermarkets, movie theatres, restaurants, school/college etc.) in easily accessible range, eventually reducing transportation requirements	2
7		**Optimizing Cloud-Based AI SaaS**- Making cloud-based AI SaaS environmentally sustainable without any IP conflict	3
8		**AI-Based Smart Thermostat/Electric Meters** to reduce the energy consumption of the HVAC systems	3
9		**Personal tracking of greenhouse gas emissions**	3

10		**Linky Meters:** Linky is a communicating or “smart” meter that transmits consumption data and receives orders remotely.	3
11		**Conscious recycling:** make the recycling supply chain accountable and consumers informed and engaged	3
12	Future of Work	**“Smart cameras for ethical surveillance/monitoring”** surroundings to avoid hazards	2
13		**Online Job Portal** – Matching job seeker skills with the relevant skills required for the job	3
14		**AI for Supply Chain Optimization** improving the efficiency of business	3
15		**AI for labour modelling** for framework development	3
16		**Training of the workforce** - Keeping a track of necessary skill sets as per job requirements using AI	3
17		**Exploring the space of working from the office v/s working from home →** AI Addressing Work-From-Home issues of efficiency	3
18		**AI-Enabled Voice Assistant** supports the workforce by creating & suggesting best practices	3
19		**AI to enhance the productivity of knowledge labour** - A spin-off company “Happiness Planet”, was developed by an R&D team at Hitachi, has been trying to observe their employee's behaviour and plans to improve their productivity by 50 times.	3
20		**AI-Enabled Smart Mobility** - Reducing the use of private vehicles and promoting sustainable modes of transport like cycling	3
21		**India Stack Program Initiative** - Part of Digital India's movement	3