



Towards Irish language capability in Generative AI

OPEN MARKET CONSULTATION DOCUMENT

Open Market Consultation for the future development of Irish language
capability in Generative AI

October 2024

Disclaimer and Copyright

All rights reserved. No part of this publication may be reproduced, stored in an automated database, or made public, in any form or by any means, electronic, mechanical, photocopying, recording or any other way, without prior written permission. This document and the accompanying annex are exclusively intended for use within the framework of and for the duration of the present market consultations within the framework of the Towards Irish language capability in Generative AI project. Any other use is not permitted, except with the prior written permission of the contracting entity. Rights of third parties may be vested in this document (including the accompanying annex).

This document (including the accompanying annex) has been drafted with the utmost care, but no guarantees are given regarding its soundness and/or completeness. Any errors or inaccuracies can be reported via email to eolas@udaras.ie

The public buyer is not responsible for the correct operation of any URL mentioned in this document, nor for the proper functioning of any used electronic platform (for example the EU survey system). Any problems encountered when using a URL and / or an electronic platform must be reported to the organisation that makes the URL or the electronic platform available. Problems with downloading and uploading (of documents) must also be reported via email to eolas@udaras.ie.

A Prior information notice, or PIN, has been published in TED to announce the Open Market Consultation on potential future procurement activity (notice publication number: eolas@udaras.ie)

The original language of this open market consultation is English.

Note: This project has been selected to benefit from assistance under the [SPIN4EIC](#) initiative to conduct this Open Market Consultation.

Abbreviations and Acronyms

AI	Artificial Intelligence
CET	Central European Time
COTS	Commercial Off-The-Shelf
EAFIP	European Assistance for Innovation Procurement
EC	European Commission
EU	European Union
FAIR	Findable, Accessible, Interoperable and Reusable
FRAND	Fair, Reasonable and Non-Discriminatory
GDPR	General Data Protection Regulation
GPA	Government Procurement Agreement
HE	Horizon Europe
IPRs	Intellectual Property Rights
OMC	Open Market Consultation
PCP	Pre-Commercial Procurement
PPI	Public Procurement of Innovative Solutions
PIN	Prior Information Notice
R&D	Research and Development
RFI	Request For Information
SMEs	Small and Medium Enterprises
SOTA	State Of The Art
TED	Tenders Electronic Daily
TRL	Technology Readiness Level
WTO	World Trade Organisation



Key Definitions

Contractor	A company or entity that has been awarded a contract under the PPI.
Public Buyer	A public entity (in this case Údarás na Gaeltachta) who purchases goods or services from the market and is subjected to the public procurement regulation.
Technology provider	A company or entity who develops and/or sells technology in the market.

Table of contents

1. Purpose of the Open Market Consultation	6
1.1. Scope and main objectives	6
1.2. Who can participate?	7
1.3. Activities & timetable	7
1.4. Registration.....	8
1.5. Procedure	8
1.6. Annexes.....	9
2. Towards Irish Language Capability in Generative AI project	10
2.1. Context and objectives	10
2.2. The Public Procurement of Innovative Solutions	10
Annex I: State-of-the-art analysis: preliminary results	12
Annex II: Request for Information	16
Request for Information questionnaire	17
Annex III: Use Case	19

1. Purpose of the Open Market Consultation

1.1. Scope and main objectives

This document describes the objectives and rules applicable to the Open Market Consultation (OMC) in preparation for the future development of Irish language capability in Generative AI technologies.

The OMC begins on the date of the publication of the Prior Information Notice (PIN) in the Tenders Electronic Daily (TED) portal and ends on the date indicated in the timetable of this document, unless Údarás na Gaeltachta decides to terminate on an early date.

Through this OMC, Údarás na Gaeltachta aims to inform the intention conduct a future public procurement process and to gather insights from the market on innovative solutions for enhancing the Irish language's presence and usability in Generative AI models. The primary focus is on developing tools that can support the Irish language in various applications, such as real-time speech processing, intelligent query responses, translation, and cultural context preservation. Additionally, the system will aim to integrate existing language corpora and datasets to ensure a comprehensive approach to Irish language technology development. This initiative seeks to bridge the gap in available resources for the Irish language in AI applications, contributing to the broader goals outlined in the Digital Plan for the Irish language (<https://www.gov.ie/en/publication/a880c-plean-digiteach/#>). Compliance with regulatory frameworks, including data protection and cybersecurity standards, is a key consideration in this project with the following required functionalities:

- Replying intelligently in Irish language to queries made in any language.
- Translating, proofreading and writing to transform and adapt Irish text in various ways.
- Ensuring that both Irish language, culture, context and experience are reflected in responses in information search engines.
- Responding in Irish to speech in real-time through chatbots and automated tools.
- Modelling Irish language across languages.
- Reflecting the complexity of the world through the lens of the Irish experience.
- Integrating existing language corpora, datasets and tools.
- Translating pictures, videos, maps.

Údarás na Gaeltachta wants to understand the technology providers' capabilities to satisfy the public buyer's needs and to obtain their input on the viability of the procurement plans and conditions as described in this document and annexes.

In sum, the objectives of this OMC are to:

- 1- Validate the findings of the State-Of-The-Art (SOTA) analysis and the viability of the set of technical and financial provisions.
- 2- Raise awareness of the industry and relevant stakeholders regarding the upcoming PPI.
- 3- Collect insights from the industry and relevant stakeholders (including users) to finetune the tender specifications.

This OMC is performed under the law of Údarás na Gaeltachta as the main public buyer, which is Irish law.

Údarás na Gaeltachta and any other contracting authorities involved in Towards Irish language capability in Generative AI project are not legally bound in any way by the outcome of the OMC.

Starting an OMC does not mean that the public buyer will start a tendering or purchasing procedure. If this OMC is followed by a tendering procedure and/or purchasing procedure, the public buyer reserves the right to adjust and/or supplement the solution described in this document on every element. No rights can be derived from statements and/or communications during this OMC in any future tendering procedure and/or purchasing procedure.

The OMC is not part of any pre-qualification or selection process. No advantage or disadvantage will be given to any technology provider / group of technology providers to the detriment of others during the OMC and the sub-sequent competitive procedure for the award of contracts.

All information provided during the OMC and other background information will be published online in English.

Where appropriate, parts of the information received from market parties can be shared with the EC.

1.2. Who can participate?

This OMC targets technology providers and other potential end users. All interested parties are invited to take part in the OMC.

Participation in the OMC is voluntary and non-binding and is at the own expense and risk of market operators. A market operator cannot charge any costs to the public buyer for the participation in the OMC and/or the (re-)use of its information in the context of a future procurement procedure.

Participation in this OMC is not a condition for submitting a tender in the subsequent procurement, does not lead to any rights or privileges for the participants, and is not part of any pre-qualification or selection process. The input provided in this OMC will not be used to evaluate future proposals.

1.3. Activities & timetable

The OMC will take place in the form of:

- An online event (in English and Irish).
- A Request for Information (RFI) – a questionnaire using the EU Survey tool.
- Other activities as deemed necessary within the scope of the project.

The timetable of the OMC activities and required actions is as follows:

Date	Event
9/10/24	Publication of the Prior Information Notice (PIN) on TED.
9/10/24	Publication of the OMC documents and EU Survey questionnaire: www.udaras.ie/ai/
14/11/24	OMC Event in English and Irish (online) 10:30-12.00 GMT. (TBC)
06/12/24	Deadline for the submission of questions via the OMC questionnaire 17:00 GMT.
20/01/24	Publication of the OMC Report. (TBC)
31/01/25	Closure of the OMC.

Table 1: OMC Timetable

The public buyer is entitled to adjust the planned activities and the timetable above and to include new activities at any time according to the needs and responses of the market. Furthermore, it may decide to terminate the OMC for its own reasons at any time. In that case, the public buyer will publish such modifications or termination on the www.udaras.ie/ai/ (and in the EU Survey: https://ec.europa.eu/eusurvey/runner/Irish_language_generative_AI).

The event and webinars celebrated within the framework of the OMC could be recorded. In that case, by attending the event, you will consent to be recorded. By using your video and microphone during the webinars you will consent to be recorded. If you do not want your voice and image to be recorded during the webinars, you may ask your questions using the chat. The public buyer shall use those records for the purpose of the project only.

In addition, please be aware that photos may be taken during the meetings. The public buyer shall use those photos for the purpose of the project only.

1.4. Registration

Parties interested in participating in the OMC activities are requested to register here: www.udaras.ie/ai/

1.5. Procedure

The OMC starts on the date of its publication in TED and ends on the date set in the timetable, unless terminated earlier.

Interested parties are requested to register through the link provided above in order to participate in the events and receive additional information about the project. The questionnaire should be filled out before the deadline indicated in the timetable above.

The public buyer will support interested parties throughout the whole OMC during the events and by answering questions through a Q&A document which will be published on the www.udaras.ie/ai/.

Additional written contributions in the form of a Request For Information (RFI) questionnaire or other questionnaires (via the EU Survey platform) aiming to collect market information on innovative and commercial solutions may be requested.

The responses to the questionnaires should not contain any confidential information. As the questionnaire is intended to explore the market “as is”, there are no wrong or right answers. The answers provided will be used as input for the procurement strategy and contract conditions.

After processing and analysing the answers, the public buyer will disseminate the results to the widest possible audience. Nevertheless, all answers provided by market parties will be anonymized and treated as confidential. The public buyer will therefore not provide information about specific answers from market operators. Only the general findings and a summary of the answers will be provided. The results of this OMC will be published on the project’s website.

In case the information provided in this document and annexes needs further clarification, market operators may ask questions during the events, or via the contact email address eolas@ударas.ie.

Market operators that wish to provide additional confidential information during the OMC can send an email to the email address indicated above. The information must be explicitly marked as confidential. Confidential information will not be included in the OMC report.

1.6. Annexes

The following annexes are part of this document:

Annex I – State-of-the-art analysis: preliminary results

Annex II – Request for Information questionnaire (published in EU Survey)

Annex III – Use case

The annexes form an integral and inseparable part of this OMC document. In the event of any conflict between the provisions of this document and the annexes, the provisions of the OMC document shall prevail.

2. Towards Irish Language Capability in Generative AI project

2.1. Context and objectives

The project "Towards Irish Language Capability in Generative AI" aims to develop and enhance the use of the Irish language within advanced AI technologies. This initiative is part of a broader strategy to integrate Irish language resources into cutting-edge AI applications, ensuring that the language can be used effectively across various digital platforms. . The primary focus of the project is on developing AI tools that support the Irish language in real-time speech processing, intelligent query responses, translation, and the preservation of cultural context within AI-generated content. Additionally, the project seeks to model the Irish language across multiple languages, reflecting its unique cultural and linguistic characteristics

To achieve these goals, the project will build on existing Irish language corpora, such as those provided by [Grúpa taighde Gaois | Fiontar & Scoil na Gaeilge, DCU](#); [Irish Language Technology Resource Marks Growth with Rebrand · ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology \(adaptcentre.ie\)](#); [ABAIR](#) archives. The project will also involve creating new datasets and tools that can support the development of AI technologies capable of handling the Irish language in a nuanced and contextually appropriate manner.

Moreover, the project aims to align with regulatory standards and protect privacy by ensuring that all AI tools developed adhere to cybersecurity protocols and personal data protection regulations, including full compliance with GDPR guidelines.

Údarás na Gaeltachta plan collaborate with the Office of the Government Chief Information Officer (OGCIO) to utilize their Generative AI Framework, which includes AI building blocks and services. The framework aims to align technology strategy with returns from key AI use cases in government, offering blueprints for a scalable, ethical, and transparent AI platform. Its core principles or pillars are to be open source, multi-model, multi-cloud, and hybrid deployment. A LLM translation service is expected to be launched under this framework which may be utilised in this project.

2.2. The Public Procurement of Innovative Solutions

This OMC concerns a future envisaged Public Procurement of Innovative solutions that integrate advanced Irish language capabilities into Generative AI technologies, enabling real-time speech processing, intelligent query responses, translation, and culturally relevant content generation.

PPI allows contracting authorities to act as a launch customer or early adopter of innovative goods or services which are not yet available on a large-scale commercial basis or require customization to the specific needs and operational environment and may include conformance testing. PPI is a specific approach for procuring innovative solutions in which procure announce well in advance the intention to buy a significant volume of innovative solutions, in order to trigger industry to bring to the market solutions with desired quality / price ratios within a specific time. Market readiness prior to deployment can be verified through e.g. conformance testing, certification or quality labelling of solutions.

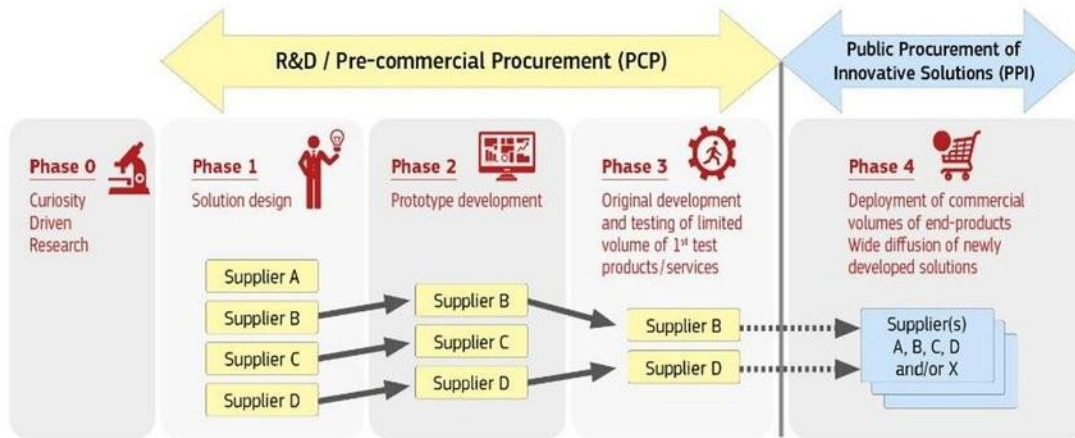


Figure 1: PCP and PPI, according to the European Commission (2016).

Based on “Pre-commercial procurement: driving innovation to ensure sustainable high quality public services in Europe”, COM(2007) 799 final.

PPI focuses on innovative solutions which are not yet available on a large-scale commercial basis. This also includes solutions based on existing technologies that are used in a new, innovative way. The solutions may have been (partially) demonstrated with success on a small scale (e.g. field testing of a first batch of products) and may be nearly or already available in small quantity on the market. However, due to residual risk or market uncertainty, the innovations are not being produced at a large scale yet and do not meet market price/quality requirements of procurers for wide deployment yet. Therefore, PPI aims to incentivize the market to deploy these solutions customized to the specific needs of the end-users.

2.3. Public Buyer: Údarás na Gaeltachta

Údarás na Gaeltachta is the regional authority established in 1980, dedicated to the economic, social, and cultural development of the Gaeltacht regions in Ireland. Its primary mission is to preserve and promote the Irish language as the main communal language of these areas. Údarás na Gaeltachta achieves this by fostering enterprise development, job creation, and supporting strategic language, cultural, and community-based activities. The organization provides a range of financial and non-financial incentives to encourage investment and innovation in various sectors, including life sciences, ICT, tourism, renewable energy, and more. For technology providers, Údarás na Gaeltachta is an important partner in advancing projects that enhance public service delivery through innovative solutions, particularly in the future development of Irish language capability.

Annex I: State-of-the-art analysis: preliminary results

This section presents the preliminary result of the market analysis and, in particular, the state-of-the-art (SOTA) analysis. The objective of this analysis was to identify existing technologies that can tackle the procurement challenge together with an analysis of the related patents and standards, and to estimate the TRL thereof.

The SOTA analysis performed within the Towards Irish Language Capability in Generative AI project reveals a complex landscape.

The results based on the patents and standards search using keywords on the IPlytics intelligent platform (<https://www.iplytics.com/>) show that there is ongoing research in the fields of natural language processing (NLP), real-time translation, and conversational AI, specifically tailored to support the Irish language.

The results, based on a thorough search of patents and standards using the IPlytics intelligent platform, indicate ongoing research and development in key areas that are critical to our use case. These areas include AI-driven real-time translation systems, advancements in NLP technologies, and innovations in conversational AI. The analysis revealed over 1,700 relevant patents, with 54 closely aligned to our use case, of which 19 were shortlisted for their direct relevance.

These technologies primarily revolve around the use of neural networks, AI-driven translation processes, and sophisticated chatbot systems. Key components include AI-driven translation systems utilizing cameras and sensors for high-accuracy text capture, optimized translation processes that enhance semantic understanding in code-switched communication, and conversational AI technologies that leverage transformer models like GPT for enhanced language understanding and response generation.

Additionally, the analysis revealed innovations like intelligent chatbots using Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) for generating natural responses, along with advanced NLP techniques that translate natural language queries into database management languages such as SQL. These technologies collectively enhance the effectiveness and efficiency of intelligent chatbots, real-time translation, and context-aware responses in the Irish language.

Examples of solutions in the market:

- Google's Dialogflow is a comprehensive platform for developing chatbots, voice bots, and virtual agents using natural language understanding and Google AI.
- Microsoft Azure offers a range of AI and NLP capabilities, including translation, speech recognition, and text analytics.
- Amazon AWS provides AI services such as Amazon Comprehend for NLP, Amazon Lex for chatbots, and Amazon Translate.
- IBM Watson offers AI and NLP solutions for various industries, focusing on enterprise applications.

- Voysis specializes in voice AI and provides tools for voice recognition, natural language understanding, and real-time speech interaction. Their platform can handle multiple languages, including Irish, making it suitable for real-time speech responses and intelligent querying.
- KantanMT offers customizable machine translation solutions that support various languages, including Irish. Their platform can be used for translation, proofreading, and adapting text while maintaining cultural and contextual accuracy.
- ICONIC specializes in neural machine translation for specific industries, ensuring high-quality translations that reflect cultural and contextual nuances.

Altocloud focuses on customer engagement and uses AI to enhance interactions, including chatbots that can handle multilingual support and real-time speech responses. Overall, there has been a lot of work done in the last years in creating standards to guarantee quality, interoperability, safety, and ethical considerations as a result of the quick development of Generative AI and NLP as well as the other fields that our use case covers.

The rapid advancement of these technologies' standards landscape can be attributed to their increasing inclusion into daily applications and ongoing improvements.

This could indicate that the field is evolving towards greater maturity, reliability, and ethical responsibility, which can ultimately drive more widespread adoption and innovation.

The following tables show the standards.

Title	Publication Date	Standard Document ID	Abstract
Language resource management — Semantic annotation framework (SemAF)	2021	ISO 24617-11	This document covers the measurable or magnitudinal aspect of quantity so that it can focus on the technical or practical use of measurements in IR (information retrieval), QA (question answering), TS (text summarization), and other NLP (natural language processing) applications
Information technology — User interfaces — Voice commands	2016	ISO/IEC 30122	ISO/IEC 30122-1:2016 defines a framework and general guidance for essential voice commands.
Language resource management — Semantic annotation framework (SemAF)	2012	ISO 24617-1	SemAF-Time, specifies a formalized XML-based markup language called ISO-TimeML, with a systematic way to extract and represent temporal information, as well as to facilitate the exchange of temporal information, both between operational language processing systems and

			between different temporal representation schemes.
Artificial Intelligence — Evaluation methods for accurate natural language processing systems	Under development	ISO/IEC AWI 23282	This document specifies the evaluation of natural language processing systems, in the sense of measuring the quality of a system's results to assess its functional suitability.
Artificial intelligence — Overview of AI tasks and functionalities related to natural language processing	Under development	ISO/IEC AWI TR 23281	This document describes the concept of AI task as applied to natural language. It proposes a landscaping of the AI tasks related to the analysis or generation of natural language, as well as other language related functionalities that are associated to those AI systems.
Information technology — Artificial intelligence — Management system	2023	ISO/IEC 42001	ISO/IEC 42001 is an international standard that specifies requirements for establishing, implementing, maintaining, and continually improving an Artificial Intelligence Management System (AIMS) within organizations.
Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)	2022	ISO/IEC 23053	This document establishes an Artificial Intelligence (AI) and Machine Learning (ML) framework for describing a generic AI system using ML technology.
Language code		ISO 639	ISO 639, Code for individual languages and language groups, can be applied across many types of organization and situations. It is needed for the basic settings of billions of user interfaces to ICT systems and devices, as well as for the indication of different language versions of websites.
Information technology — Coding of audio-visual objects	2019	ISO/IEC 14496-3	This document integrates many different types of audio coding: natural sound with synthetic sound, low bitrate delivery with high-quality delivery, speech with music, complex soundtracks with simple ones, and

			traditional content with interactive and virtual-reality content.
Translation services — Requirements for translation services	2015	ISO 17100	ISO 17100:2015 provides requirements for the core processes, resources, and other aspects necessary for the delivery of a quality translation service that meets applicable specifications.
Translation services — Post-editing of machine translation output — Requirements	2017	ISO 18587	ISO 18587:2017 provides requirements for the process of full, human post-editing of machine translation output and post-editors' competences.

Table 2: Standarts

Conclusions: The SOTA analysis reveals that while there is a significant array of advanced NLP and AI technologies available in the market that can support our use case, capturing the unique cultural and contextual nuances of the Irish language may require additional customizations. The technologies identified provide a solid foundation for real-time translation, intelligent chatbot development, and natural language processing in Irish. However, specific needs like reflecting Irish culture and experience, proofreading, and translating visual content might necessitate further specialized tools and innovations beyond the current state of the art.

Keywords used for the patents and standard searches

1. Generative AI
2. Irish language
3. Natural language processing
4. Search engines
5. Real-time speech
6. Translation
7. Queries
8. Chatbot

Annex II: Request for Information

The Request for Information survey is part of the OMC of the Towards Irish Language Capability in Generative AI project. It should provide the public buyer with market feedback on solutions that support a more efficient integration of the Irish language in AI systems, incorporating privacy and security by design.

Technology providers are invited to answer all the questions of the survey (one survey per company). The results will be considered when drafting the tender documents for the future PPI.

The survey should be filled out online and submitted via the following link: https://ec.europa.eu/eusurvey/runner/Irish_language_generative_AI

Please note that taking part in this survey is not a prerequisite for participation in the future PPI and does not give any advantage to any technology provider. All information provided in the questionnaire will be anonymized, summarized and published online in English on www.udas.ie/ai

Your personal data will be collected, processed, stored and used by the public buyer with the only purpose of gathering information from the market within the framework of Towards Irish Language Capability in Generative AI project. Personal data will be treated as strictly confidential according to the General Data Protection Regulation (Regulation 2016/679 of the European Parliament and of the Council - GDPR). You may exercise your right to access your personal data and the right to rectify such data by contacting: eolas@udas.ie.

Request for Information questionnaire

QUESTIONS FOR TECHNOLOGY PROVIDERS	
Procurement challenge and requirements	
1	Do you have any suggestions regarding the scope of the envisaged procurement?
2	If you were to test and customise relevant solutions, could you indicate an estimated budget for this? Please justify your answer.
3	Do you have knowledge of any suitable technology or combination of technologies for supporting the Irish language in AI applications? Yes / No. If yes, please elaborate.
4	Do you know any developments in the field of natural language processing and AI-driven translation for minority languages which we need to take into account? Yes / No. If yes, which ones?
5	Do you foresee any barriers to implement a solution for real-time Irish language translation and AI conversational agents? Yes / No / I do not know. If yes, please elaborate
6	Can you tackle all or part of the requirements of this challenge? Yes / No / I do not know yet. If yes, please explain. If I do not know yet, what additional information would you need?
7	Can you identify relevant functionalities that have not been described in the market consultation document? Yes / No If yes, please elaborate.
8	Can you provide any other recommendations regarding the integration of Irish language capabilities into AI systems?
State-of-the-art (SOTA) analysis	
9	In your assessment, do you believe the current state of technology adequately addresses the needs outlined in the procurement? Please explain.
10	What existing solutions or methodologies do you propose for effectively meeting the requirements of the procurement?
11	Can you provide information on the current Technology Readiness Level (TRL) of the proposed solutions or methodologies? How mature are these solutions in terms of practical implementation and deployment?

12	Are there any existing patents or industry standards that you believe are pertinent the development of AI solutions for the Irish language? Please specify any relevant intellectual property or regulatory frameworks.
13	Are there any existing patents that may present challenges or barriers to delivering a solution in the context of the envisaged procurement? Please elaborate on any potential patent-related obstacles.
Miscellaneous	
14	What information do you still need in order to make a good plan of action for the development and/or implementation of solutions suitable to address Irish language processing in AI?
15	Do you have specific requirements or questions about project on the Towards Irish Language Capability in Generative AI? Yes / No. If yes, which ones?
16	How could you contribute to providing innovative and regulatory compliant solutions for integrating Irish language capabilities into AI systems? Please explain.
17	What are the risks associated to the development and implementation of a solution that tackles the functional needs of the project?
18	Do you have any suggestions and/or remarks?

Annex III: Use Case

Towards Irish language capability in Generative AI

Introduction

The project is contextualized in the framework of the national policy set in the Digital Plan for the Irish language ([gov - Plean Digiteach \(www.gov.ie\)](#)) to broaden the use and applicability of a number of shared language corpora predominantly structured text datasets ([Grúpa taighde Gaois | Fiontar & Scoil na Gaeilge, DCU](#); [Irish Language Technology Resource Marks Growth with Rebrand · ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology \(adaptcentre.ie\)](#); [ABAIR](#)).

Ireland is one of twelve EU countries in the ‘*Alliance for Language Technologies – European Digital Innovation Consortium*’ (‘*ALT-EDIC*’). This seeks to address the lack of training data for official languages of the EU.

Údarás na Gaeltachta has undertaken a consultation and mapping exercise across the civil and public service to identify ongoing work in this area. This has identified further corpora and language resources, such as the video, text and sound archive of TG4, RTÉ and broadcast archive of RTÉ RnaG.

This collaboration allows these shared corpora to become the cornerstone of generative AI enabled Irish language and large language models fine-tuned to reflect the Irish experience through history and as a significant cultural influence across the globe.

Tools based on these corpora using existing Generative AI capability (such as GPT) would allow for improved public services as Gaeilge, in real time, across a number of domains. It will allow those currently working on basic services as Gaeilge to add greater value and possibly address more challenging and interesting work. Údarás seeks to build on emerging capacity to build an ecosystem of not only Irish language resources but low resource language models that can meet emerging international demand for similar services across virtually all languages outside English.

Based on the initial information, some foreseen functionalities and use case are detailed in the following sections.

Functionalities:

- Replying intelligently in Irish language to queries made in any language.
- Translating, proofreading and writing to transform and adapt Irish text in various ways.
- Ensuring that both Irish language, culture, context and experience are reflected in responses in information search engines.
- Responding in Irish to speech in real-time through chatbots and automated tools.
- Modelling Irish language across languages.
- Reflecting the complexity of the world through the lens of the Irish experience.
- Integrating existing language corpora, datasets and tools.
- Translating pictures, videos, maps.

Keywords: Generative AI, Irish language, natural language processing, search engines, real-time speech, translation, queries, chatbot, GPT.

“As is”	Desired situation
Describe how it is done now	Describe how you wish it to be
<p>Use case 1: Written and verbal queries End-user: citizens, public officials E.g. User has a question about garbage collection calendar</p>	<p>Written and/or speech reply in accurate Irish language with optional pictures, calendars and instructions. Functionality: writing and replying to queries Datasets:</p>
<p>Use case 2: Translation to Irish language E.g. User needs to translate an official document.</p>	<p>A document in any language can be translated and the Irish translation is proofread and accurate in context. Functionality: translate, proofread, contextualize Datasets:</p>
<p>Use case 3: Irish search engines E.g. User inserts a question or keywords to find results.</p>	<p>Search engine provides results in Irish language Functionality: modelling search scenarios Datasets:</p>
<p>Use case 4: Irish real-time speech E.g. User asks (Siri) to provide the weather forecast</p>	<p>AI assistant and/or Chatbot responds back in speaking in Irish language Functionality: Datasets:</p>