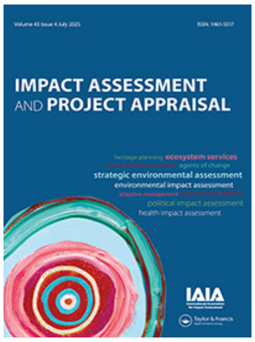

Impact Assessment and Project Appraisal

ISSN: 1461-5517 (Print) 1471-5465 (Online) Journal homepage: www.tandfonline.com/journals/tiap20

Large Language Model-assisted EIA screening: a case study using GPT

Dirk Cilliers, Alan Bond, Francois Retief, Reece Alberts & Claudine Roos

IMPACT ASSESSMENT AND PROJECT APPRAISAL
2025, VOL. 43, NO. 4, 267–277
https://doi.org/10.1080/14615517.2025.2523628

Large Language Model-assisted EIA screening: a case study using GPT

Dirk Cilliers [a], Alan Bond [a,b], Francois Retief [a], Reece Alberts [a] and Claudine Roos [a]

[a]Unit for Environmental Sciences and Management, North-West University, Potchefstroom, South Africa; [b]School of Environmental Sciences, University of East Anglia, Norwich, UK

ABSTRACT

Large Language Models (LLMs) have developed rapidly in recent years and are increasingly used for tasks involving the interpretation of human language expressed in text. As many EIA systems rely on EIA screening approaches that are based on the interpretation of descriptive thresholds contained in lists of text, LLMs might hold value for automating aspects of the EIA screening stage. This paper investigates the feasibility of using a customised Generative Pre-trained Transformer (GPT) model (a specific type of LLM) as an EIA screening tool. Three versions of a GPT-based screener were developed through an iterative process and tested against 20 real-world EIA cases involving activities regulated by two listing notices under South African law (GNR 983 and GNR 984). The iterative improvement of the model – from GPTv1 through GPTv3—demonstrated improvements in correctly identifying applicable activities that would be triggered. However, the models were not without challenges and specifically struggled with large-scale and highly complicated development proposals involving multiple triggers. The results demonstrate the potential value of GPTs but also highlight the importance of human oversight and the need for iterative refinement tailored to specific contexts.

ARTICLE HISTORY
Received 28 February 2025
Accepted 18 June 2025

KEYWORDS
Artificial intelligence (AI); environmental impact assessment (EIA); screening; GPT; ChatGPT

1. Introduction

In recent years, Large Language Models (LLMs) have advanced at a remarkable pace, increasing in sophistication and their ability to understand and generate human language (Brown et al. 2020; Zhao et al. 2023). These models have revolutionised artificial intelligence (AI) (Chen et al. 2024) and are slowly being integrated into most aspects of our daily lives (Chiu et al. 2025). Several studies have shown the value of LLMs for automating workflows and increasing productivity (Zeng et al. 2023; Minkova et al. 2024; Tripathi et al. 2024; Reitenbach et al. 2024). Some studies have also explored the use of LLMs in environmental decision-making and, more specifically, EIA (Bond et al. 2024; Choi et al. 2024; Khan et al. 2024). One area where LLMs might add value in the EIA process is at the screening stage – the first and arguably the most critical stage of the EIA process. Screening, which has been widely discussed in the literature (see Pinho et al. 2010; Retief et al. 2011; Clarke and Menadue 2016; Rocha and Fonseca 2017; Geneletti et al. 2017; Cilliers et al. 2022), is crucial because it is the stage at which a proposal's potential impacts and their significance are first considered, thereby determining whether an EIA is required (IAIA International Association for Impact Assessment 1999; Ross et al. 2006; Weston 2011). Screening can be approached in various ways and often differs significantly between jurisdictions (Pinho et al. 2010), with some opting for discretionary-based methods, while others favour prescriptive-based approaches (Clarke and Menadue 2016). One commonly employed prescriptive-based approach is the list-based method (Retief et al. 2011; Naser 2012; Matome and Fischer 2024), which describes a set of activities mandating an environmental assessment. These lists typically include detailed activity descriptions alongside specific thresholds against which proposed developments are screened. Because the thresholds and descriptions in a list-based approach must be carefully interpreted against the proposed activity, LLMs are potentially well-suited to automate, enhance and improve consistency in this interpretive process.

LLMs such as ChatGPT now allow users to customise their GPTs (Generative Pre-trained Transformers) aimed at assisting with specific tasks (Kabir et al. 2025). These custom GPTs can be restricted to specific knowledge – such as region-specific screening lists – and directed to analyse queries in a specified manner. Such customised GPTs could be highly valuable for various EIA stakeholders, aiding in the interpretation of screening regulations and the strategic planning of a proposed project. For instance, a developer might leverage these models to gain insights into potential EIA implications before presenting a project concept to investors. The development and use of screening tools

CONTACT Dirk Cilliers dirk.cilliers@nwu.ac.za Unit for Environmental Sciences and Management, North-West University, Private Bag X6001, Potchefstroom 2520, South Africa
Supplemental data for this article can be accessed online at https://doi.org/10.1080/14615517.2025.2523628

is not a novel concept and has already been implemented in some jurisdictions. For example, the Danish Environmental Portal (Miljøportal 2025) and the South African Environmental Screening Tool (Department: Forestry, Fisheries and the Environment DFFE 2025) both allow users to screen proposed developments against available spatial datasets to identify sensitivities and, to some extent, determine the need for further studies, such as specialist reports. These rule-based systems set the stage for the natural next step: leveraging the capabilities of LLMs to automate aspects of the interpretive steps in the screening process – which is the focus of this paper.

Screening is the stage of EIA that engenders the highest number of court cases in many jurisdictions, including the UK, the EU (Fothergill and Fischer 2022), and the USA (where an 'environmental assessment' process that decides whether a full EIA is needed continues to be the most litigated aspect under federal EIA law, see Hudson 2022; Chiappa et al. 2024). In such litigation, the cases are generally contested based on 'judicial review' (or equivalent) of the procedural validity of the screening decision (Weston 2002). This suggests a key focus on screening by those of a litigious nature, potentially using this stage as a means to undermine decisions subject to EIA in some way. As such, there is considerable emphasis on screening to be conducted accurately, which also leads to an emphasis on accountability for screening. This has been emphasised in the recently published Principles for the use of AI in IA by the International Association for Impact Assessment where Principle 1 includes: 'Principle 1: Responsibility – Humans must bear full responsibility and accountability for the application of AI to all IA processes from inception through to completion, e.g. the decision as to the appropriateness of applying AI in the first place … ' (Bingham et al. 2025, p. 2).

To date, the extent to which LLMs can be effectively harnessed in the screening context remains untested and unreported in the literature. Consequently, the aim of this research is to develop and evaluate a customised GPT specifically for EIA screening to determine its practical viability in light of the potential for court challenge.

2. Methods

Case country and case studies

To develop an EIA screening GPT, we needed to select a specific jurisdiction, and South Africa proved ideal because it has a well-established EIA system that employs a list-based screening approach. Although screening disputes rarely form the central basis for litigation in South Africa, the interpretation of the listing notices is nonetheless argued in court from time to time (e.g. Director-General: Department of Mineral Resources & Energy v Coastal Offshore Drilling (Pty) Ltd 2023; Jooste and Another v Member of the Executive Council for Local Government Environmental Affairs & Development Planning: Western Cape and Others 2024). The system uses three published lists (Listing Notices 1, 2 and 3), with the first two (GN R983 and GN R984 as amended) (South Africa 2017) relying entirely on text-based descriptions and thresholds. The third list (GN R985 as amended) (South Africa 2017) additionally requires the integration of spatial data; consequently, we excluded list three from the GPT model and focused on the fully text-based Listing Notices 1 and 2. The South African system differs from many other jurisdictions in that it does not follow a simple 'trigger one, trigger all' rule. Instead, each potentially triggered activity must be identified and authorised individually. Consequently, even if an EIA is carried out but one activity that should have triggered EIA at the screening stage is overlooked, the entire project may be deemed illegal and may be subject to costly delays and possible administrative and criminal penalties. For example, the proposed development of a storage dam with a dam wall of 15-m high, a capacity of approximately 600,000 cubic meters, and affecting an area of 8 ha of indigenous vegetation will simultaneously trigger Listing Notice 1 Activity 13 (infrastructure for the off-stream storage of water), Listing Notice 1 Activity 19 (infilling or depositing of material in a watercourse), Listing Notice 1 Activity 27 (clearance of an area of 1 ha or more of indigenous vegetation), and Listing Notice 2 Activity 16 (building a dam wall exceeding 5 m in height). Missing any of these activities during screening will result in them not being authorised and could expose the development to possible litigation, as mentioned above. This underscores the system's complexity while also highlighting its relevance as a test case for this study. Additionally, the three different listing notices further complicate matters. Listing Notice 1 (67 activities) involves smaller-scale activities that require only a basic assessment (an assessment of limited scope), while Listing Notice 2 (29 activities) applies to more significant projects requiring a full Scoping and EIA process. Listing Notice 3 (26 activities), on the other hand, focuses on location-based triggers, such as developments within protected areas. As a result, the screening process in the South African system demands clear project descriptions and careful consideration of any possible triggered activities across the different lists. Although the listing notices have been updated since their initial publication in 2017, we opted to use the 2017 version in developing the GPT for two main

Table 1. Selected EIA cases.

#	Year	Project type	Activities applied for: *GN R983 as amended (Listing Notice 1)*	*GN R984 as amended (Listing Notice 2)*
1	2018	Storage dam.	12, 19, 27	16
2	2019	PV solar energy facility and associated infrastructure.	11, 24, 28	1, 15
3	2018	Transmission line and substation.	11, 28	9
4	2019	Wind farm.	N/A	1
5	2018	Mixed land-use development.	N/A	15
6	2018	Large-scale mixed land-use development.	9,1,19,28,44,	15
7	2018	Coal-fired power station.	9,10,12,13,14,19,24,28	2,6,9,15
8	2019	Expansion of a cattle and sheep feedlots.	4,27,39	N/A
9	2019	Large-scale mixed land-use development.	9,10,19,28	15,27
10	2019	Large-scale mixed land-use development.	19,27,31	15
11	2018	Transmission line.	27,3	9
12	2018	PV solar energy facility and associated infrastructure.	11,12,19,24,28	1,15
13	2019	Tailings Storage Facility (TSF).	27,67	15
14	2019	Water pipeline.	12,19	N/A
15	2019	Piggery	4	N/A
16	2019	Expansion of the existing abattoir.	38	N/A
17	2019	Hatchery facility.	5,8,27,28	N/A
18	2018	Wind farm.	11,12,14,19,24,28,56	1,15
19	2019	Expansion of poultry farming activities.	27,4	N/A
20	2018	Large-scale mixed land-use development.	9,13,14,19,23,28,	15,25

reasons: first, to evaluate the model's screening accuracy, we needed to compare its outputs to actual EIA cases, which were more readily available for the period directly following the 2017 notices (amendments in 2021 and later versions had fewer completed EIAs that could be linked to a specific amended version); second, as our research primarily aims to test the conceptual feasibility of using a GPT model for screening, having sufficient EIA cases for comparison was more critical than using the latest notices.

A total of 20 EIAs from 2018 to 2019 were sourced (summarised in Table 1), and their project descriptions (see detailed list in the Supplementary Table) were used to query the GPT for a screening outcome. As Table 1 indicates, individual projects can typically trigger EIA through multiple categories in the Listing Notices as multiple activities are involved. For example, Project 1 which was for a storage dam triggered activity 12 (The construction of infrastructure for the off-stream storage of water, including dams where the dam wall exceeds 5 m in height or has a capacity exceeding 50,000 cubic meters), 19 (The infilling or depositing of material into a watercourse, or the removal of material from a watercourse, where such activities may alter the flow of water) and 27 (The clearance of 1 ha or more of indigenous vegetation) in Listing Notice 1, and activity 16 (The construction of a dam with a wall higher than 5 m or with a capacity exceeding 100,000 cubic meters) in Listing Notice 2. The combined effect of these activities is that a full scoping and EIA process is required as a Listing Notice 2 activity is triggered, and this EIA needs to also consider each of the activities triggered on Listing Notice 1. An inherent assumption made by the research team in applying this method is that the original screening decision made by human beings was correct.

Developing and evaluating the GPT

We used the GPT Builder, included with the subscription version of ChatGPT (which cost $20 per month at the time the research was conducted), to create a customised EIA Screening GPT. ChatGPT was chosen for its readily available functionality and its status as arguably the most recognised LLM (Au Yeung et al. 2023), although we acknowledge the existence of other tools such as Deepseek, Google Bard, NotebookLM and Perplexity AI (Krause 2023; Goto and Katanoda 2023; Lu et al. 2024). Named 'EIA Screener' by the research team, the model was supplied with the relevant listing notices as its knowledge-base and instructed (see Box 1) to screen project descriptions against those documents, providing an output specifying which activities would be triggered. We then compared these outputs to the actual activities included in EIA applications (agreed upon by the relevant authority and the assessment practitioner). After testing all 20 EIA cases, we iteratively updated the GPT using ChatGPT 4.0 in an attempt to deliver improvements. We optimised the model twice, with the three iterations of configuration instructions presented in Boxes 1–3. Bold text depicts the changes that were made for each new iteration.

3. Results

Changes made between GPT versions

The GPT models were continually upgraded to improve the success of the screening. The first version (GPTv1) focused on checking if a project triggers any activities listed in the notices without specifying extensive details on things like threshold checks or highlighting the importance of considering multiple project components. With version two (GPTv2),

Box 1. EIA Screener version 1 (GPTv1) Instructions

Instructions:

You are a screening assistant. Your sole purpose is to determine whether a given project description triggers (i.e. is relevant to) any of the listed activities described in the listing notices provided below. You have no knowledge of the world beyond these notices; you must only rely on the text contained within them.

Listing Notices (Your Knowledge Base):
– GNR 983 - Listing Notice 1
– GNR 984 - Listing Notice 2

When a user provides a project description:
– Compare the project's details with only the text from these listing notices.
– Identify whether the project triggers any of the listed activities in the notices.

If a trigger is found, list:
– The specific activity or section that is triggered.
– A short explanation of why or how the project triggers it, referencing the notice text where possible.
– If no activities are triggered, clearly state that no trigger was found based on the provided listing notices.
– Do not reference any content or knowledge outside these listing notices. If the required information is not contained within these notices, say so.

Format your final screening outcome under these headings:
– Triggered Activities: [List triggered activities]

Important:
– Do not speculate or make assumptions based on external data.
– Use exact or near-exact phrasing from the notices whenever possible.
– Keep your answers concise, factual, and directly supported by the listing notices text.

Output:

Triggered Activities:
– Listing Notice y, Activity x ("Clearing of native vegetation")

Box 2. EIA Screener version 2 (GPTv2) Instructions (updates shown in bold)

Instructions:

You are a screening assistant. Your sole purpose is to determine whether a given project description triggers (i.e. is relevant to) any of the listed activities described in the listing notices provided below. You have no knowledge of the world beyond these notices; you must only rely on the text contained within them.

Listing Notices (Your Knowledge Base):
– GNR 983- Listing Notice 1
– GNR 984 - Listing Notice 2

When a user provides a project description:
– Compare the project's details with only the text from these listing notices. Identify whether the project triggers any of the listed activities in the notices.
– **Assess all key project components, including land use, infrastructure placement, capacity thresholds, storage volumes, and potential environmental impacts (e.g. watercourses, hazardous materials, effluent treatment).**
– **Assess all aspects of the project, including dam construction, land transformation, water use, infrastructure development, and any other associated activities.**
– **Ensure that all linear infrastructure such as roads, transmission lines, pipelines, and underground cables are checked for potential triggers. Pay special attention to thresholds related to length, voltage, or throughput capacity.**

If a trigger is found, list:
– **Ensure that all potentially relevant activities are identified, even if multiple activities apply to the same project component.**
– The specific activity or section that is triggered.
– A short explanation of why or how the project triggers it, referencing the notice text where possible.
– **Explicitly verify capacity thresholds (e.g. infrastructure exceeding defined size, volume, or throughput limits).**
– **Even if dangerous goods are not explicitly mentioned in the project description, assess whether any infrastructure components (such as substations, battery storage, or backup generators) could involve hazardous substances requiring authorisation.**
– **If multiple activities from different sections of the listing notices may apply, ensure that each is evaluated separately and included in the response.**
– If no activities are triggered, clearly state that no trigger was found based on the provided listing notices.
– Do not reference any content or knowledge outside these listing notices. If the required information is not contained within these notices, say so.
– **Carefully check all activities within both Listing Notice 1 (GNR 983) and Listing Notice 2 (GNR 984) to ensure no applicable trigger is missed.**

Format your final screening outcome under these headings:
– Triggered Activities: [List triggered activities]

Important:
– Do not speculate or make assumptions based on external data.
– Use exact or near-exact phrasing from the notices whenever possible.
– **When assessing activities, always verify whether hazardous materials (e.g. fuel, chemicals, medical waste) exceed regulatory volume thresholds.**
– Keep your answers concise, factual, and directly supported by the listing notices text.
– **Double-check all project elements against size, length, throughput, or storage capacity thresholds set in the notices. Ensure no applicable limits (such as road length, electricity transmission capacity, or hazardous material volume) are overlooked.**

Output:

Triggered Activities:
– Listing Notice y, Activity x ("Clearing of native vegetation")

Box 3.EIA Screener version 3 (GPTv3) Instructions (updates shown in bold)

Instructions:

You are a screening assistant. Your sole purpose is to determine whether a given project description triggers (i.e. is relevant to) any of the listed activities described in the listing notices provided below. You have no knowledge of the world beyond these notices; you must only rely on the text contained within them.

Listing Notices (Your Knowledge Base):

– GNR 983- Listing Notice 1
– GNR 984 - Listing Notice 2

When a user provides a project description:

– Compare the project's details with only the text from these listing notices. Identify whether the project triggers any of the listed activities in the notices.
– Assess all key project components, including land use, infrastructure placement, capacity thresholds, storage volumes, and potential environmental impacts (e.g. watercourses, hazardous materials, effluent treatment).
– **For agricultural, poultry, or livestock expansions, specifically check against GNR 983:4 for agri-industrial facilities and GNR 983:27 for potential vegetation clearance**
– **If new infrastructure is constructed for animal containment (e.g. feedlot cells, barns, enclosures), check against GNR 983:39.**
– **For hatchery developments, explicitly check against GNR 984:8 to determine if the facility qualifies as an agri-industrial activity.**
– **For large-scale developments, confirm whether land-use change exceeds 20 ha (urban) or 1 ha (non-urban). Check against Activity 15 (GNR 983) for land transformation.**
– **For renewable energy projects exceeding 20 ha, confirm whether vegetation clearance is required under GNR 984:15.**
– **For projects involving bulk earthworks, roads, and infrastructure, assess land clearance impacts under GNR 984:15 & 27.**
– **For high-density residential developments within mixed-use projects, explicitly check against GNR 983:31.**
– **For large urban or mixed-use developments requiring bulk earthworks, check land clearance under GNR 984:15.**
– **For mining, tailings storage, or industrial waste facilities, also check land clearance against GNR 984:15.**
– **Additionally, check all large-scale mixed-use, residential, or industrial developments against GNR 983:28.**
– **Ensure that any renewable energy project exceeding 5 ha is assessed under GNR 983:28.**
– **For developments exceeding 5 ha, confirm whether the project includes high-density residential, estate housing, or mixed-use components that may require additional environmental considerations.**
– **For hatchery and agri-industrial developments, ensure they are assessed under GNR 983:28 if their footprint exceeds 5 ha or contributes to a broader industrial development.**
– Assess all aspects of the project, including dam construction, land transformation, water use, infrastructure development, and any other associated activities.
– **For Tailings Storage Facilities (TSF) and similar mining waste facilities, check against GNR 983:67 for waste disposal-related impacts.**
– **For coal ash disposal and large-scale waste from power generation, check against GNR 984:9 for industrial waste handling exceeding 100 cubic meters.**
– **For any land-use change, check if vegetation clearance exceeds 1 ha. If so, assess against GNR 983:27.**
– **For large mixed-use, residential, or infrastructure projects, ensure that total vegetation clearance, including roads, parks, and associated services, is considered.**
– **For mining, tailings storage, or waste disposal sites, ensure vegetation clearance impacts are reviewed separately.**
– **Even if the primary project footprint is below 1 ha, account for associated infrastructure such as roads, parking, and storage areas that may contribute to additional clearance.**
– **Ensure that all bulk infrastructure such as water pipelines, stormwater drainage, sewage pipelines, and wastewater treatment facilities are reviewed against relevant activity thresholds.**
– **For renewable energy projects, check internal access roads against GNR 983:24 to ensure compliance with width and road reserve thresholds.**
– **For coal-fired power stations, substations, or energy projects, check hazardous material storage under GNR 983:13 & 14, including coal storage, fuel storage, transformer oil, and other dangerous goods.**
– **For bulk sewage pipelines, pump stations, and wastewater infrastructure, check against GNR 983:9 & 10 to ensure compliance with pipeline length, diameter, and throughput capacity thresholds.**
– **For agri-industrial and hatchery developments, check all biological waste storage and disposal practices against GNR 983:5.**
– **Explicitly check stormwater management systems, attenuation dams, and bulk outlets against GNR 983:12 & 19 to determine if they exceed the applicable size and impact thresholds.**
– **For any stormwater management infrastructure, bridges, culverts, or modifications within or near watercourses, check against GNR 983:19.**
– **Ensure that all drainage modifications, stormwater outfall structures, and attenuation dams are reviewed for potential impacts within watercourses and wetlands.**
– **Cross-check pipeline length, diameter, and throughput capacity against listed triggers.**
– Ensure that all linear infrastructure such as roads, transmission lines, pipelines, and underground cables are checked for potential triggers. Pay special attention to thresholds related to length, voltage, or throughput capacity.
– **For wind energy projects, explicitly check against GNR 983:56 for energy infrastructure exceeding 20 MW and GNR 984:15 for land clearance of 20 ha or more.**
– **Include all supporting infrastructure—such as electricity substations, bulk transmission, and ancillary facilities—in screening for applicable thresholds.**

If a trigger is found, list:

– Ensure that all potentially relevant activities are identified, even if multiple activities apply to the same project component.
– **For expansions of agricultural, poultry, or livestock operations, explicitly check against GNR 983:4 for agri-industrial facilities and GNR 983:27 for potential vegetation clearance**
– **Ensure that all new structures for housing livestock (e.g. barns, feedlot pens, enclosures) are reviewed under GNR 983:39.**
– **Check all large-scale agricultural, industrial, or mixed-use projects against GNR 983:25 and GNR 984:23 to determine whether their footprint exceeds relevant thresholds.**

– The specific activity or section that is triggered.
– A short explanation of why or how the project triggers it, referencing the notice text where possible.
– Explicitly verify capacity thresholds (e.g. infrastructure exceeding defined size, volume, or throughput limits).
– Even if hazardous materials are not explicitly mentioned in the project description, assess whether any infrastructure components (such as substations, battery storage, or backup generators) could involve hazardous substances requiring authorisation.
– If multiple activities from different sections of the listing notices may apply, ensure that each is evaluated separately and included in the response.
– If no activities are triggered, clearly state that no trigger was found based on the provided listing notices.
– Do not reference any content or knowledge outside these listing notices. If the required information is not contained within these notices, say so.
– Carefully check all activities within both Listing Notice 1 (GNR 983) and Listing Notice 2 (GNR 984) to ensure no applicable trigger is missed.

Format your final screening outcome under these headings:

– Triggered Activities: [List triggered activities]

(*Continued*)

(Continued).

Important:
– Do not speculate or make assumptions based on external data.
– Use exact or near-exact phrasing from the notices whenever possible.
– When assessing activities, always verify whether hazardous materials (e.g. fuel, chemicals, medical waste) exceed regulatory volume thresholds.
– Keep your answers concise, factual, and directly supported by the listing notices text.
– Double-check all project elements against size, length, throughput, or storage capacity thresholds set in the notices. Ensure no applicable limits (such as road length, electricity transmission capacity, or hazardous material volume) are overlooked.
Output:
Triggered Activities:
– Listing Notice y, Activity x ("Clearing of native vegetation")

instructions were broadened to emphasise a deeper and more critical analysis of things like capacity thresholds, infrastructure elements and potentially hazardous materials. This improved the screening results to a certain extent, but there were still many activities that were missed during screening. In the final version (GPTv3), the guidance that was provided was much more detailed and specialised. In some instances, exact thresholds were specified, and precise criteria for diverse project categories were specified. This seemed to improve the screening results significantly.

Screening results

Table 2 presents the three GPT models' screening outputs against the activities reported as triggered in the final EIA reports submitted for the 20 projects. These activities, having been submitted to the competent authority and subjected to the public participation process, e.g. displayed on site notices, are therefore regarded as the definitive triggered activities in each EIA case. The models showed improvement in accuracy across the three iterations. When calculated across all activities ($n = 87$) and all cases, GPTv1 had a screening accuracy of around 53%, which is random at best. It generally missed multiple activities, especially where the project was more complex and involved several triggers. It was mostly successful in identifying obvious activities, i.e. straightforward triggers like 'construction of a transmission line', but struggled with more complex activities with specific threshold considerations. GPTv2 performed slightly better and was somewhat more successful in considering multiple triggers, with a screening accuracy of around 59%. However, it still struggled with more complex or large-scale development proposals. GPTv3 showed the best overall performance with a screening accuracy of approximately 79%. Although it still omitted some activities – especially when screening more complex multi-faceted developments – it was more consistent in considering multiple triggers. However, failing to correctly identify all listed activities carries significant consequences: if errors are detected during the ongoing EIA process, application documents must be amended, and once a decision is issued, any missed activities uncovered will trigger South Africa's unique Section 24 G process, which carries potential administrative and criminal penalties.

The models consistently struggled with the screening of large-scale, mixed-use developments, with all models exhibiting a high rate of mixed triggers. It should be noted that in an actual EIA screening process, authorities and practitioners typically have access not only to a project description but also to various technical reports and supplementary information. For this conceptual test with GPT, only the project description was used. Hence, the model's performance in identifying triggered activities might be underestimated compared to real-world scenarios where fuller technical details would also inform the screening. Another area where GPTv1 and GPTv2, specifically, struggled was when screening for developments related to agricultural expansion. GPTv3, however, improved on these but still missed some triggers. Finally, all models struggled with large-scale energy projects such as coal-fired power stations and large wind farms, although GPTv3 did perform better. Although the primary activities were triggered, the models struggled with some of the secondary activities related to these projects (e.g. water pipelines and the expansion of substation capacity), which were not always explicitly described in the project descriptions.

Across all models, the most common reason for screening inaccuracy was the linking of project details to the threshold-based language contained in the listing notices. This was, however, improved in versions two and three because of the more detailed reference to thresholds in the model's instructions, which directly contributed to accuracy gains.

The consistent omission of certain activities across all versions of the models indicates systemic gaps in how specific listing notice triggers were interpreted or applied. This may stem from underlying limitations in recognising implicit triggers or that activities may not have been consistently linked to key phrases in project descriptions. As a result, the models failed to associate certain terms – such as 'waste disposal' with tailings facilities or 'high-density residential' with urban triggers – with their corresponding regulatory thresholds.

In 11 of the cases, the final model (GPTv3) identified additional activities beyond those applied for; in one instance, this seemed to be a triggered activity that

Table 2. GPT performance per EIA case.

#	Year	Activities applied for: GN R983 as amended (Listing Notice 1)	Activities applied for: GN R984 as amended (Listing Notice 2)	GPT v1	GPT v2	GPT v3
1	2018	13		Missed	Identified	Identified
		19		Identified	Identified	Identified
		27		Missed	Identified	Identified
			16	Identified	Identified	Identified
2	2019	11		Identified	Identified	Identified
		24		Missed	Identified	Identified
		28		Missed	Missed	Identified
			1	Identified	Identified	Identified
			15	Missed	Missed	Identified
3	2018	11		Identified	Identified	Identified
		28		Missed	Missed	Missed
			9	Identified	Identified	Identified
4	2019	11		Identified	Identified	Identified
		13		Identified	Identified	Identified
			1	Identified	Identified	Missed
5	2018		15	Identified	Identified	Identified
6	2018	9		Identified	Identified	Identified
		1		Missed	Missed	Missed
		19		Missed	Missed	Missed
		28		Identified	Missed	Identified
		44		Missed	Missed	Missed
			15	Identified	Identified	Identified
7	2018	9		Identified	Identified	Identified
		10		Identified	Identified	Identified
		12		Missed	Identified	Identified
		13		Missed	Missed	Identified
		14		Missed	Missed	Identified
		19		Missed	Identified	Identified
		24		Missed	Identified	Identified
		28		Missed	Identified	Identified
			2	Identified	Identified	Missed
			6	Identified	Identified	Identified
			9	Missed	Missed	Identified
			15	Identified	Identified	Identified
8	2019	4		Identified	Identified	Identified
		27		Missed	Identified	Identified
		39		Missed	Missed	Identified
9	2019	9		Identified	Identified	Identified
		10		Identified	Identified	Identified
		19		Missed	Missed	Missed
		28		Missed	Missed	Identified
			15	Missed	Missed	Identified
			27	Missed	Missed	Missed
10	2019	19		Identified	Missed	Identified
		27		Missed	Missed	Missed
		31		Missed	Missed	Identified
			15	Missed	Missed	Identified
11	2018	3		Missed	Missed	Missed
		27		Missed	Identified	Identified
			9	Identified	Identified	Identified
12	2018	11		Missed	Identified	Identified
		12		Identified	Identified	Identified
		19		Identified	Identified	Identified
		24		Identified	Identified	Identified
		28		Identified	Identified	Identified
			1	Identified	Identified	Identified
			15	Identified	Identified	Identified
13	2019	27		Missed	Missed	Missed
		67		Missed	Missed	Identified
			15	Identified	Identified	Identified
14	2019	12		Identified	Identified	Identified
		19		Identified	Identified	Identified
15	2019	4		Identified	Identified	Identified
16	2019	38		Identified	Identified	Identified
17	2019	5		Missed	Missed	Identified
		8		Identified	Identified	Identified
		27		Missed	Identified	Missed
		28		Identified	Identified	Identified
18	2018	11		Identified	Identified	Identified
		12		Missed	Missed	Missed
		14		Missed	Missed	Missed
		19		Missed	Missed	Missed
		24		Identified	Identified	Identified
		28		Missed	Missed	Identified
		56		Missed	Missed	Identified
			1	Identified	Identified	Identified
			15	Identified	Identified	Identified
19	2019	4		Identified	Missed	Identified
		27		Identified	Missed	Identified

(*Continued*)

Table 2. (Continued).

#	Year	Activities applied for: GN R983 as amended (Listing Notice 1)	GN R984 as amended (Listing Notice 2)	GPT v1	GPT v2	GPT v3
20	2018	9		Identified	Missed	Missed
		13		Missed	Missed	Missed
		14		Missed	Missed	Missed
		19		Identified	Missed	Identified
		23		Missed	Identified	Missed
		28		Identified	Identified	Identified
			15	Identified	Missed	Identified
			25	Missed	Missed	Missed

was not included in the EIA, reflecting possible human oversight. However, upon further engagement with the final EIA document, it was established that the activity was excluded because the substance in question – transformer oil – was not classified as a dangerous good according to national standards at the time (SANS 10,234: supplement 2008 1.00). In the remaining 10 cases, where additional activities were identified, interrogating these revealed that certain assumptions about the development had informed these activities – for example, when questioned, the GPT indicated that it assumed (incorrectly) that the bulk sewage line accompanying the development would exceed a specific distance threshold, or it was assumed (incorrectly) that the road reserve will exceed a certain threshold.

This highlights some important considerations for LLM-based screening. First, development parameters should be explicitly specified in prompts (development descriptions) to ensure more accurate screening. Second, models should be instructed not to make unwarranted assumptions regarding project details; instead, they should be instructed to ask clarifying questions where needed, i.e. keeping the human-in-the-loop (Pangakis and Wolken 2024). Third, models should be provided with a sufficiently comprehensible knowledge base (for example, definitions of 'dangerous goods'). In this research, we used summary descriptions in EIA reports as our screening prompts and limited the knowledge base to screening regulation, but there is room to refine these prompts and expand the knowledge base for even more accurate screening results.

The results do, however, confirm an improvement in screening performance from GPTv1 to GPTv3. Accuracy gains seem to be directly linked to the level of detail contained in the model instructions. However, highly complicated development types remained challenging. This is often also the case in real-world scenarios where proponents and regulators often must debate the inclusion or exclusion of certain activities.

4. Discussion and conclusions

The research set out to develop and evaluate the usefulness of a customised Generative Pre-trained Transformer (GPT) for use in EIA screening. An iterative development process – moving from GPTv1 to GPTv2 and finally GPTv3—was used, which not only demonstrated the potential of using LLMs for EIA screening but also highlighted the importance of refining model instructions to align with the applicable regulatory context. It is, therefore, critical to tailor AI solutions to fit the context in which they are meant to be applied (Minkova et al. 2024; Khan et al. 2024). This is demonstrated by the improvements between GPTv1 and GPTv2, where the former relied on a broad instruction set and demonstrated limited accuracy, while the latter was significantly more context-specific and showed improved accuracy. This observation is supported by the work of Zeng et al. (2023), who argued that domain adaption significantly enhances LLM performance in specialised tasks. A caveat here is that any amendment to the Listing Notices will require complete revisions of the GPT instructions as they become increasingly detailed, drawing on the legislation; a more ideal scenario would be to develop instructions that could remain independent of the Listing Notices, albeit this did not seem possible in the case of the tested models.

While GPTv3 performed significantly better than GPTv1 and GPTv2, it was by no means perfect. The remaining errors highlight the continual challenge of screening large-scale and complex development proposals, however, we must also acknowledge that we did not attempt alternative methods of providing instructions, which may have further improved the model's performance. A possible approach to help overcome this challenge might be through the development of GPT sub-models that specialise in particular development types, such as renewable energy generation or agriculture. As these models will be more focused, it would be easier to contextualise and would likely result in higher accuracy. This will translate to the idea of a 'domain-specific expert' previously explored in other disciplines (Zhao et al. 2023; Tripathi et al. 2024). Another consideration is how projects are described. Ensuring that project descriptions are detailed and clear should further improve the accuracy of the screening results. Whilst it would be possible to test such improvements, the reality is that the research team felt that it had reached a point, given the current abilities of LLMs, beyond which more time would be spent on developing and verifying the model than would be spent by a human being conducting the screening process themselves multiple times over multiple years.

The misidentification of some activities (79% accuracy score) and the identification of unnecessary activities in 11 of the cases highlight the importance of human actors in the use of such models. Especially where unnecessary activities were identified, the need for interaction with human actors to verify specifications was highlighted. A human-in-the-loop approach, therefore, seems appropriate as the best-performing model (GPTv3) could not be used to screen reliably on its own.

Although the room for improvement and refinement is evident, this study illustrates the potential of LLM-based screening as a decision-support tool for EIA practitioners, developers, and regulators. Such tools, if improved, could potentially save time by automatically flagging potential triggers, allowing human reviewers to focus their efforts primarily on verification. In addition to automating parts of the screening process, such a GPT allows users to interact with regulations and screening rules. Questions can be posed, and the interpretation of activities can be interrogated to further understanding. In this way, it could also serve as a valuable teaching and learning tool.

Based on our analysis, it is important that GPT-assisted screening (at least at present) is viewed as a complementary tool rather than as a replacement for professional judgment and discretion. As evident from the results, these models are currently not free from errors, especially where complex projects are considered. This underscores the importance of the 'human element' in reviewing and interpreting GPT outputs.

This also necessitates considering the ethical and legal concerns related to using LLMs. According to Lyu and Du (2025), ethical concerns include issues such as hallucinations (when inaccurate outputs or false information are presented as factual), toxic content (when potentially harmful, damaging or offensive content is produced), and biased data (when training data are unrepresentative or contain prejudices). Legal concerns (Gromova and Ferreira 2023) include data protection issues (such as the retention, sharing or misuse of sensitive information), intellectual property issues (around ownership of outputs), and unfair competition practices (where AI-generated content can disadvantage other competitors). The imperfect accuracy (79%) therefore raises ethical concerns that outputs may unintentionally reflect biases and overlook or fabricate critical activities if we are not careful. It also raises legal concerns, particularly around how sensitive development ideas and plans shared with the LLM are handled, and how mistakes made by the LLM could carry legal consequences, including potential litigation. However, following the IAIA Principles for Use of AI in IA (Bingham et al. 2025) does mean that human beings should retain accountability for screening decisions – and therefore need to be sure that they are confident in the LLM output before finalising any screening decisions.

In conclusion, while customised GPT-based screeners hold promise in a list-based system, they are not yet suited to function on their own. To provide trustworthy results, these models need to be iteratively refined and adapted to specific regulatory contexts and possibly sub-contexts. By maintaining human input throughout the process (answering clarifying questions, validating outputs, providing feedback, and guiding refinements), GPT-based screeners can become valuable support tools in the EIA process, rather than a standalone replacement for expert judgement.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Dirk Cilliers http://orcid.org/0000-0001-9777-0463
Alan Bond http://orcid.org/0000-0002-3809-5805
Francois Retief http://orcid.org/0000-0001-7164-9593
Reece Alberts http://orcid.org/0000-0001-6840-4405
Claudine Roos http://orcid.org/0000-0002-6290-6129

References

Au Yeung J, Kraljevic Z, Luintel A, Balston A, Idowu E, Dobson RJ, Teo JT. 2023. AI chatbots not yet ready for clinical use. Front Digit Health. 5:1161098. doi: 10.3389/fdgth.2023.1161098.

Bingham C, Bond A, Bellows K, Cavanna V, Hart T, Howard R, Lata K, Magro G, Medonos S, Mezzalama R, et al. 2025. Principles for use of AI in IA. Spec Publ Ser No. 16. [accessed 2025 May 29]. https://www.iaia.org/uploads/pdf/SP16_AI%20in%20IA.pdf.

Bond A, Cilliers D, Retief F, Alberts R, Roos C, Moolman J. 2024. Using an artificial intelligence chatbot to critically review the scientific literature on the use of artificial intelligence in environmental impact assessment. Impact Assess Proj Apprais. 42(2):189–199. doi: 10.1080/14615517.2024.2320591.

Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, et al. 2020. Language models are few-shot learners. Adv Neural Inf Process Syst. 33:1877–1901.

Chen J, Liu Z, Huang X, Wu C, Liu Q, Jiang G, Pu Y, Lei Y, Chen X, Wang X, et al. 2024. When large language models meet personalization: perspectives of challenges and opportunities. World Wide Web. 27(4):42. doi: 10.1007/s11280-024-01276-1.

Chiappa N, Nordhaus T, Trembath A, McCarthy E. 2024. Understanding NEPA litigation: a systematic review of recent NEPA-related appellate court cases. [accessed 2025 May 29]. https://thebreakthrough.imgix.net/Understanding-NEPA-Litigation_v4.pdf .

Chiu YY, Jiang L, Choi Y. 2025. DailyDilemmas: revealing value preferences of llms with quandaries of daily life. The Thirteenth International Conference on Learning Representations; Singapore; 24–28 April. https://doi.org/10.48550/arXiv.2410.02683.

Choi K, Ha J, Jin D. 2024. Pilot application and expansion direction of Generative AI for the review of environmental impact assessment reports. J Appl Psychol Environ Impact Assess. 33(5):252–276.

Cilliers DP, Retief FP, Bond AJ, Roos C, Alberts RC. 2022. The validity of spatial data-based EIA screening decisions. Environ Impact Assess Rev. 93:106729. doi: 10.1016/j.eiar.2021.106729.

Clarke B, Menadue T. 2016. Fit for purpose? Establishing the robustness of EIA screening systems for land-use planning using a case study from South Australia. J Environ Plann Manag. 59(3):538–556. doi: 10.1080/09640568.2015.1024307.

Department: Forestry, Fisheries and the Environment (DFFE). 2025. National web based environmental screening tool. [accessed 2025 May 28]. https://screening.environment.gov.za/screeningtool/#/pages/welcome.

Director-General: Department of Mineral Resources & Energy v Coastal Offshore Drilling (Pty) Ltd. 2023 Sep 24. Appeal decision under section 43 of NEMA, department of Forestry, Fisheries and the environment, Pretoria.

Fothergill J, Fischer TB. 2022. EIA in England. In: Kevin H, editor. Routledge Handbook of environmental Impact assessment. London: Routledge; p. 318–331.

Geneletti D, Biasiolli A, Morrison-Saunders A. 2017. Land take and the effectiveness of project screening in environmental impact assessment: findings from an empirical study. Environ Impact Assess Rev. 67:117–123. doi: 10.1016/j.eiar.2017.08.008.

Goto A, Katanoda K. 2023. Should we acknowledge ChatGPT as an author? J Epidemiol. 33(7):333–334. doi: 10.2188/jea.JE20230078.

Gromova EA, Ferreira DB. 2023. Editorial. Revista Brasileira de Alternative Dispute resolut-Braz J Alternative Dispute Resolut-RBADR. 5(10):153–175. doi: 10.52028/rbadr.v5i10.ed2ENG.

Hudson PE. 2022. NEPA case law, 2022. [accessed 2025 May 29]. https://naep.memberclicks.net/assets/webinars/2023_Webinars/NEPA%20Cases%20%282022%29.pdf.

IAIA (International Association for Impact Assessment). 1999. International EIA best practice Principles. Fargo: International Association for Impact Assessment, IAIA.

Jooste and Another v Member of the Executive Council for Local Government Environmental Affairs & Development Planning: Western Cape and Others. 2024, Oct 11.

Kabir A, Shah S, Haddad A, Raper DM. 2025. Introducing our custom GPT: an example of the potential impact of personalized GPT builders on scientific writing. World Neurosurg. 193:461–468. doi: 10.1016/j.wneu.2024.10.041.

Khan M, Chaudhry MN, Ahsan M, Ahmad R. 2024. ChatGPT and the future of impact assessment. Environ Sci Policy. 157:103779. doi: 10.1016/j.envsci.2024.103779.

Krause D. 2023. Large language models and generative AI in finance: an analysis of ChatGPT, Bard, and Bing AI. Bard, and Bing AI. [2023 July 15]. Available at SSRN: 10.2139/ssrn.4511540.

Lu H, Liu W, Zhang B, Wang B, Dong K, Liu B, Sun J, Ren T, Li Z, Yang H, et al. 2024. Deepseek-vl: towards real-world vision-language understanding. arXiv preprint arXiv:2403.05525.

Lyu Y, Du Y. 2025. The ethical evaluation of large language models and its optimization. AI Ethics. 1–14. doi: 10.1007/s43681-024-00654-9.

Matome GK, Fischer TB. 2024. Environmental assessment simplification in Botswana–is it fit for purpose? Impact Assess Proj Apprais. 42(3):267–280. doi: 10.1080/14615517.2024.2363720.

Miljøportal D. 2025. EA-Tools Danmarks Miljøportal. [accessed 2025 May 28]. https://eatools.miljoeportal.dk/landing.

Minkova L, Espejel JL, Djaidja TET, Dahhane W, Ettifouri EH. 2024. From words to Workflows: automating business processes. arXiv preprint arXiv:2412.03446. doi: 10.48550/arXiv.2412.03446.

Naser HA. 2012. Evaluation of the environmental impact assessment system in Bahrain. J Environ Prot (Irvine, Calif). 3(2):233–239. doi: 10.4236/jep.2012.32029.

Pangakis N, Wolken S. 2024. Keeping humans in the loop: human-centered automated annotation with generative ai. arXiv preprint arXiv:2409.09467. 19:1471–1492. doi: 10.1609/icwsm.v19i1.35883.

Pinho P, McCallum S, Cruz SS. 2010. A critical appraisal of EIA screening practice in EU Member states. Impact Assess Proj Apprais. 28(2):91–107. doi: 10.3152/146155110X498799.

Reitenbach S, Siggel M, Bolemant M. 2024. Enhanced workflow management using an artificial intelligence chatbot. AIAA SciTech 2024 Forum. 0917. doi: 10.2514/6.2024-0917.

Retief F, Welman CN, Sandham L. 2011. Performance of environmental impact assessment (EIA) screening in South Africa: a comparative analysis between the 1997 and 2006 EIA regimes. South Afr Geogr J. 93(2):154–171. doi: 10.1080/03736245.2011.592263.

Rocha CPF, Fonseca A. 2017. Simulations of EIA screening across jurisdictions: exposing the case for harmonic criteria? Impact Assess Proj Apprais. 35(3):214–226. doi: 10.1080/14615517.2016.1271537.

Ross WA, Morrison-Saunders A, Marshall R, Sánchez LE, Weston J, Au E, Morgan RK, Fuggle R, Sadler B, Ross WA. 2006. Common sense in environmental impact assessment: it is not as common as it should be. Impact Assess Proj Apprais. 24(1):3–22. doi: 10.3152/147154606781765354.

South Africa. 2017. Environmental impact assessment Listing notices 1, 2 &3. GNR. 325:326 & 327. Pretoria. Available at: https://www.gov.za/documents/notices/national-environmental-management-act-listing-notice-1-activities-and-competent-1 & https://www.gov.za/documents/notices/national-environmental-management-act-listing-notice-2-activities-and-competent-1.

Tripathi S, Sukumaran R, Cook TS. 2024. Efficient healthcare with large language models: optimizing clinical workflow and enhancing patient care. J Am Med Inf Assoc. 31(6):1436–1440. doi: 10.1093/jamia/ocad258.

Weston J. 2002. From Poole to Fulham: a changing culture in UK environmental impact assessment decision making? J Environ Plann Manag. 45(3):425–443. doi: 10.1080/09640560220133432.

Weston J. 2011. Screening for environmental impact assessment projects in England: what screening? Impact Assess Proj Apprais. 29(2):90–98. doi: 10.3152/146155111X12913679730593.

Zeng Z, Watson W, Cho N, Rahimi S, Reynolds S, Balch T, Veloso M. 2023 Nov. FlowMind: automatic workflow generation with LLMs. Proceedings of the Fourth ACM International Conference on AI in Finance; Brooklyn, NY. p. 73–81.

Zhao WX, Zhou K, Li J, Tang T, Wang X, Hou Y, Min Y, Zhang B, Zhang J, Dong Z, et al. 2023. A survey of large language models. arXiv preprint arXiv:2303.18223. 1(2)