expert.ai

NLP Solution Evaluation:

POC Best Practices



Introduction

Artificial Intelligence (AI) technologies possess immense potential to revolutionize the way businesses operate by enhancing efficiency, automating processes, providing better customer support, delivering superior business insights and mitigating operating risks. However, like any new technology, introducing AI also entails certain risks. These include uncertainties about the technology's ability to deliver the anticipated business outcomes, concerns regarding its accuracy, the ability to explain and audit its behavior, and ensuring a sustainable Total Cost of Ownership.

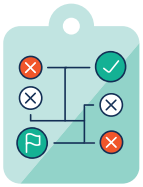
This is why a proof of concept (PoC) is such a critical step before investing in any technology or solution, but especially with AI technologies.

The Value of a Proof of Concept for AI and NLP Solutions

Regardless of where or how you use natural language processing (NLP), text analytics or any "intelligent" technology, AI is a major investment that will transform your business. And there are plenty of statistics that talk about the difficulty of bringing projects from idea to implementation. A recent Gartner executive survey revealed that around "54% of AI projects make it from pilot to production."[1]

Nonetheless, investment in AI is on the rise and businesses are looking to take advantage of a variety of capabilities for a diverse set of business and technical use cases. Our own survey found that nearly 80% of organizations surveyed expect to spend more on NLP projects in the next 12 to 18 months. Meanwhile, the frenzy over ChatGPT, for example, has prompted 45% of executives to increase their AI investment and investigation, according to a recent Gartner poll.[2] So, given the challenges and opportunities, companies need to know how to best move forward from the outset and not be caught by surprise later in the process. This is where a proof of concept can help.

So, what makes such projects so difficult? Let's look at one scenario.



Natural language processing (NLP) is one tool that organizations can use to solve problems like validating claims against trusted sources, measuring customer sentiment or extracting language entities that are needed to automated follow-on processes related to language data. However, NLP is not deterministic. It is empirical and probabilistic. Outputs can change dramatically in response to small tweaks of a model. The relationship between input and output is not always obvious. Training and tuning can vary widely depending on the framing of the problem. For example, a model that needs to classify "school buses" vs. "automobiles" can perform very well using a small training set, but a model that must accurately classify "Porsche Taycan" vs. "Porsche Panamera" will need more data points.

When evaluating NLP, it's therefore important to think carefully about the intended application of the AI model, the source inputs and the expected benefits of an implementation.

The Proof of Concept

A proof of concept helps a company validate the feasibility of the proposed solution in a controlled environment and it provides an opportunity to identify potential risks and challenges that may arise during the implementation phase.

A PoC can provide a concrete demonstration of the capability of AI technology—using real output—to solve your real problem. It's the best way to understand whether a use case can really achieve positive benefits from that approach. Essentially, it can determine if that approach will be successful or not.

Aside from the technical capabilities, it can also help determine an organization's readiness and understand if you have the necessary resources and expertise to deploy and maintain the solution.

PoCs benefit the decision-making process by allowing the business to:

- Test the technologies and methodologies to be used
- Deliver more immediate and concrete value
- Quickly compare different solutions and approaches
- Gain experience, skills and confidence in model outputs

Proof of Value (PoV) vs Proof of Concept (PoC)

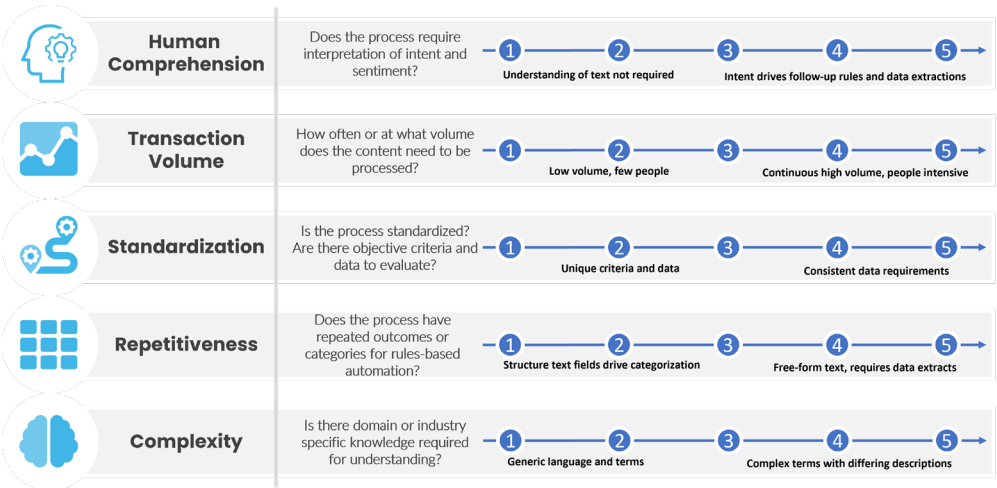
A PoC is typically used to demonstrate the technical feasibility of a concept or idea. It is a small - scale version of a solution that is created to test specific features or functionality. **Goal:** Validate that a technology or approach can be used to solve a problem.	A PoV is used to demonstrate the business value of a software solution. It is designed to show how a specific technology or approach can deliver measurable benefits to the organization. **Goal:** Prove that a solution can solve a business problem or provide a competitive advantage.

Prepare for the PoC: Best Practices

1. Start with the problem you want to solve

To get started, it's important to identify the problems you need to solve and to be sure that what you're asking of the technology is realistic. Some problems and use cases may not be well suited for AI. Some use cases have too low a volume to justify the investment in automation and AI. Other use cases may have too low a value. For additional insights on selecting the right use cases and what it takes to successfully deliver an NLP project, explore expert.ai's "The Roadmap to NLP Success" guide.

- How reliant is the process on human understanding of language?
- What is the volume and frequency of the process?
- Is the process standardized?
- Are there repeatable outcomes and categories?
- Does the process require subject matter experts (SMEs) for correct interpretation?





2. Identify the necessary skills

In addition to considering the problems you'd like to solve, you'll also want to focus on the internal skills that your teams may need. Since NLP implementations are different from other kinds of software implementations (probabilistic vs deterministic), the way you think about them and the way you go about them – *and the skills and capabilities you will need from your vendor* – will be different. Your partner vendor should be able to provide you with guidance here.

3. Define success criteria

Setting success factors for an NLP use case should consider both technical and business success. For instance, a definition for the accuracy of the model output thresholds needs to be grounded in the yield needed to provide an ROI in production and project approval. A proof of concept can also provide a secondary benefit of helping a business understand if an existing process must be replicated in the same way when using AI or if it's better to combine AI with other processes and techniques.

4. What to expect from a vendor

A PoC allows you to quickly compare different approaches and to test a vendor's capability. Vendors who promote a PoC approach will accelerate your AI-based solution benchmark, and you'll want to work with an AI technology partner that can provide both experience and guidance. Experienced vendors will offer a well-structured, yet flexible, PoC methodology and timeline that will accomplish the defined goals.

A PoC is typically a significant time and resource investment for both the vendor and customer. Following a well-considered methodology with an experienced team can help to ensure PoC process success. Specifically for enterprise text analytics use cases, Gartner recommends that data and analytics leaders should "conduct flash or quick proofs of concept to test a vendor's capability with a small sample of data to narrow down the options, then plan for a full ontology of domain-specific data for a comprehensive evaluation of the tools' capabilities."[3]

Expert.ai PoC Methodology

Expert.ai is a world leader in practical natural language processing (NLP), with an enterprise hybrid AI with 300+ production AI deployments. Focused exclusively on natural language technologies (NLT), expert.ai has participated in hundreds of NLP evaluations. The methodology described here is the product of that experience – to accurately evaluate NLP as fit for purpose.

The expert.ai NLP evaluation methodology is flexible, but typically includes these steps:

Step	Description	Ownership	Rough Timeline
Use Case Definition	Business process and assets under solution scope	Customer Provides & expert.ai Reviews	Weeks 1 and 2
Data Points for Evaluation	Define which data points will be used for evaluation and how they will be evaluated (e.g., "count of treatments" might seem clear, but if "X-Ray" appears multiple times in a document, it is important to define what counts as a unique treatment)	Customer Provides & expert.ai Reviews	
Use Case Evaluation Target	Target use case subset to prove business value before a full production implementation	Customer Provides & expert.ai Reviews	
Success Criteria Definition	Define the model output thresholds needed to yield ROI in production and project approval	Joint Creation	
Corpus Compilation	A small number (~100) of target documents that are as close to production documents as possible used for model evaluation	Customer Provides & expert.ai Reviews	Weeks 3 and 4
Corpus Training Set vs Test Set	Divide the corpus into a training set and a test set (for sets <10 docs, training and test corpus should be the same; for sets >100 docs, bias towards training, e.g., 75 training and 35 test)	expert.ai Performs	
Corpus Validation	Documents uploaded into training environment	expert.ai Performs	

Step	Description	Ownership	Rough Timeline
Model Testing	Baseline model applied to the training corpus, matching against the categorization and extraction targets, and expert.ai shares initial untuned results	expert.ai Performs & Customer Reviews	Weeks 5 and 6
Model Tuning	Different model tuning strategies (e.g., retrain an ML model with updated inputs) are employed, and expert.ai shares tuned outputs for review (Iterative step – repeated as needed)	expert.ai Performs & Customer Reviews	Weeks 7, 8 and 9
Success Criteria Review	Review of model outputs against predefined success threshold criteria	Joint Review	Week 10

PoCs typically last 6 to 12 weeks depending on their complexity, and run with a weekly rhythm, where a set of results is reviewed each week.

Using this methodology, customers get a consistent and high-quality set of evaluation data, a realistic estimate of the time and effort required for the production implementation, and the experience of real expert.ai NLP engagement, just on a smaller scale.

Here is a summary framework of the different phases and activities that you can use to prepare for your proof of concept:

Step	Description
Use Case Definition	Clearly define each use case with expected outcomes. **Example:** "Classify attorney demand packages as urgent (<10 days remaining) or non-urgent (>10 days before due)."
Data Points for Evaluation	Specific data points that will be targeted in the PoC. **Example:** "In a demand package, identify date of letter, date of receipt, due date (calculated or extracted)."
Define Success Criteria and ROI Threshold	Model output thresholds and criteria required for ROI. **Example:** "Date of letter, due date, categorize as urgent if due date is less than 10 days from current date. This will reduce the number of urgent letters missed by 20%, yielding a ROI of $1M."
Targeted Corpus	Select representative document examples with a consistent file format and resolution. **Example:** "10 real sample demand packages with a variety of demand letter formats and data points of interest (e.g., different date formats, different expressions of due date)."

Step	Description
Corpus Validation	Perform OCR quality checks. **Example:** "Confirm extraction targets exist and identify source text(s)"
Baseline Model	Run the Corpus through model and compare JSON outputs to targets and generate metrics. **Example:** "Review AI strategies (i.e., common features of demand letters such as presence of People entities with "Esq" designations for "Esquire") and initial results with prospect."
Model Tuning	Depending on initial results, reannotate, refine rules, try additional model algorithms. **Example:** "Output should include page of demand letter, values of data points for evaluation, and whether the package is 'Urgent' or 'Not Urgent'.
Success Criteria Review	Confirm that the Success Criteria has been met. **Example:** "Model output correctly categorizes as 'Urgent' all eligible demand packets. Model also correctly extracts and normalizes the Date of letter and due date 90% of the time."

Conclusion

Implementing an AI-based NLP solution doesn't have to be long and complicated. Choosing the right partner for your projects is key to ensuring that you're starting your AI journey with solid forethought, planning, expertise and best practices. This way, you'll be positioned to implement the solution best suited for your needs and place you ahead of the competition without time delays or challenging your expense budget.



[1] Gartner, "Gartner Survey Reveals 80% of Executives Think Automation Can Be Applied to Any Business Decision"
https://www.gartner.com/en/newsroom/press-releases/2022-08-22-gartner-survey-reveals-80-percent-of-executives-think-automation-can-be-applied-to-any-business-decision

[2] Gartner, "Gartner Poll Finds 45% of Executives Say ChatGPT Has Prompted an Increase in AI Investment"
https://www.gartner.com/en/newsroom/press-releases/2023-05-03-gartner-poll-finds-45-percent-of-executives-say-chatgpt-has-prompted-an-increase-in-ai-investment

[3] Gartner, "Market Guide for Text Analytics"
https://www.gartner.com/document/3989657?ref=solrAll&refval=367039397

Get Started

Contact expert.ai to learn more.



expert.ai

About us

Expert.ai (EXAI:IM) is a leading company in AI-based natural language software. Organizations in insurance, banking and finance, publishing, media and defense all rely on expert.ai to turn language into data, analyze and understand complex documents, accelerate intelligent process automation and improve decision making. Expert.ai's purpose-built natural language platform pairs simple and powerful tools with a proven hybrid AI approach that combines symbolic and machine learning to solve real-world problems and enhance business operations at speed and scale. With offices in Europe and North America, expert.ai serves global businesses such as AXA XL, Zurich Insurance Group, Generali, The Associated Press, Bloomberg INDG, BNP Paribas, Rabobank, Gannett and EBSCO.

www.expert.ai **info@expert.ai**

© 2023 Expert AI. All rights reserved.