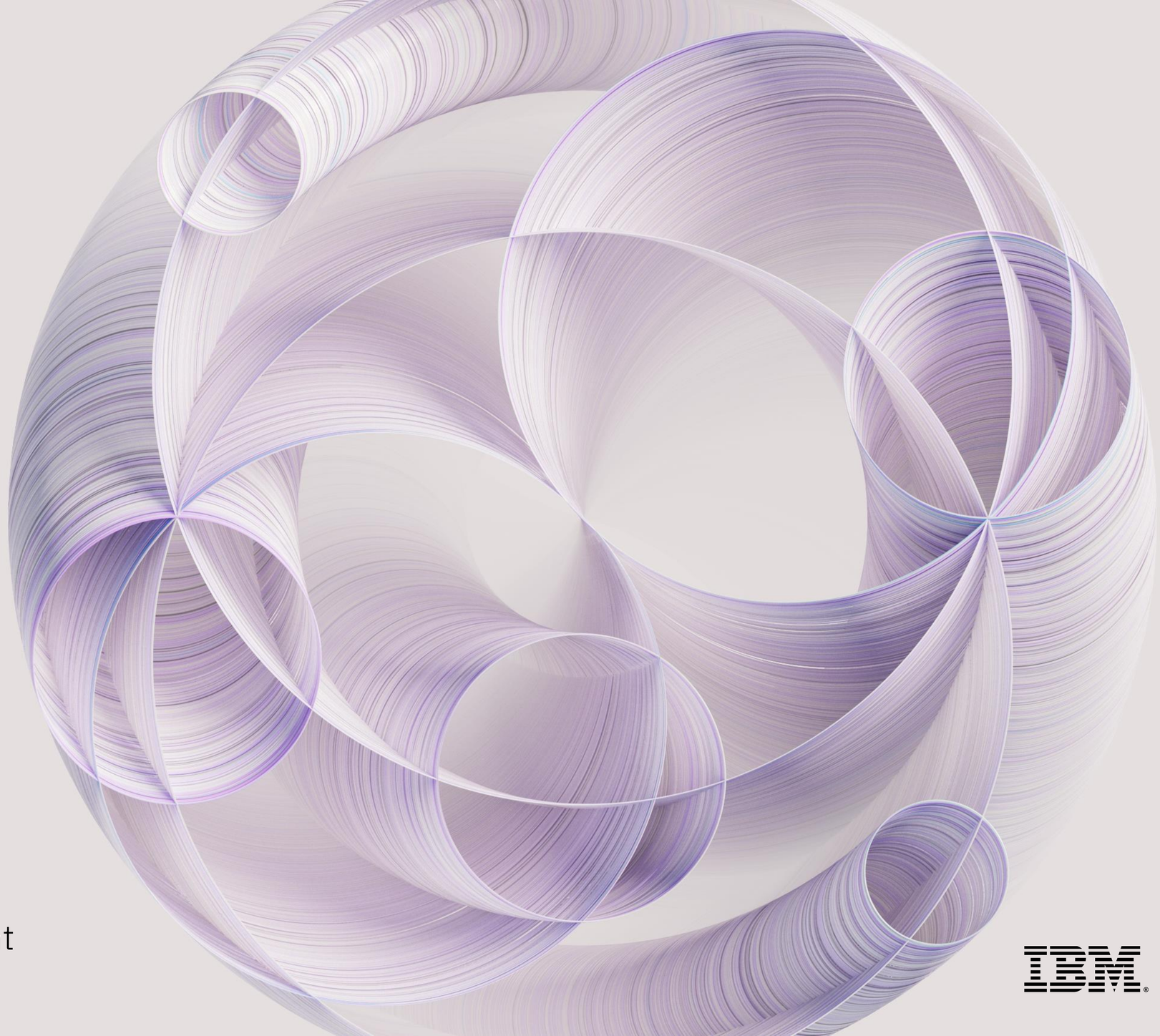
Watsonx.ai PoX education

Best practices for watsonx.ai



Felix Lee
felix@ca.ibm.com
Principal, Learning Content Development
Data and AI

IBM

Seller guidance and legal disclaimer

IBM and Business Partner
Internal Use Only

Slides in this presentation marked as **"IBM and Business Partner Internal Use Only"** are for IBM and Business Partner use and should not be shared with clients or anyone else outside of IBM or the Business Partners' company.

The information contained in this publication is provided for informational purposes only. While efforts were made to verify the completeness and accuracy of the information contained in this publication, it is provided AS IS without warranty of any kind, express or implied. In addition, this information is based on IBM's current product plans and strategy, which are subject to change by IBM without notice. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this publication or any other materials. Nothing contained in this publication is intended to, nor shall have the effect of, creating any warranties or representations from IBM or its suppliers or licensors, or altering the terms and conditions of the applicable license agreement governing the use of IBM software.

References in this presentation to IBM products, programs, or services do not imply that they will be available in all countries in which IBM operates. Product release dates and/or capabilities referenced in this presentation may change at any time at IBM's sole discretion based on market opportunities or other factors and are not intended to be a commitment to future product or feature availability in any way. Nothing contained in these materials is intended to, nor shall have the effect of, stating or implying that any activities undertaken by you will result in any specific sales, revenue growth, or other results.

All client examples described are presented as illustrations of how those clients have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by client.

Content

- Alternatives to PoX
- Best practices for PoX
  - Scope correctly based on available skills/resources
  - Only what you can deliver successfully
  - Avoid complexity
  - Show differentiation
- Planning
  - Client discovery
  - Planning
  - Document of Understanding (DoU)

Before you agree to a PoX

Offer alternatives

Alternative validation methods:

- Engage IBM TechXchange
- Pre-built demos and POTs
- Discussions with engineers or PM

Set expectations

- Client needs to participate
- Clearly defined success criteria, where a client says “We will buy if you show x, y, and z”
- Scope limited to reasonable time and effort
- Focus on OOTB functionality, standard tooling, pre-built integrations
- DOU required

Know what happens next

- Next steps in buying process
- Client commitments for successful completion
- IBM watsonx Demo Hub

Only what you can do with available resources

The principles:

- Make sure your team has the skills and availability to complete all deliverables
- Don’t count on delivery help without a funding commitment
- Minimize committing to anything you can’t do yourself

What this means in the context of watsonx.ai:

- Potential skill gaps
  1. Custom environments (non-IBM cloud)
  2. Integrating and configuring with other non-IBM products (a vector database, other applications)
  3. CP4D installation/configuration (if the client is looking for on-premises watsonx.ai)
  4. Generative AI governance questions
- **Best case:** the technical seller has the skills to deliver everything; aim for that!
- Ask the account team for help in engaging additional resources

Only what you can successfully deliver

The principles:

- Avoid ‘science projects’ and ‘moon shots’
- Don’t commit to production-level readiness
- Don’t bog down in subtle complexities of the use case

What this means in the context of watsonx.ai:

1. Decline PoX projects on esoteric, capability-stretching use cases
2. Stick to out-of-the-box functionality and configurations standard tooling, and integrations
3. Avoid committing to production levels of readiness
4. Avoid extensive “fine-tuning” model parameters
   - Use defaults;
   - When necessary, tune one at a time
5. Avoid complexities with environments

Minimal effort and complexity

The principles:

- Minimize time spent on setting up the PoX environment
- Limit to familiar models
- Limit the amount of monitoring/metrics to what is necessary to verify capabilities
- Minimize integrations with client systems – limited customer content or data that requires substantial cleansing/masking
- Show a capability once – no multiple examples of the same feature

What this means in the context of watsonx.ai:

1. Use the TechZone environment whenever possible
2. watsonx.ai rolls out new models frequently; avoid trying new ones that you have never experienced before in a PoX
3. Work with the client to settle on the minimum metrics and tests required to verify the PoX.
4. No PII, credit card info, or other confidential info
   - Use synthetic data (if applicable)
5. RAG use case – either with or without vector database, not both.

Show differentiating capabilities

The principle:

The scope must include capabilities
that differentiate IBM from other vendors,
or you have just spent weeks validating
a lower-cost competitor’s solution

What this means in the context of watsonx.ai:

1. Highlight (some of the) following watsonx.ai benefits:
   - IBM’s indemnification policies
   - Eco-system with watsonx.data and .governance
   - Hybrid capabilities
   - Curated/filtered data for IBM-developed models

2. Include (some of the) following items in the PoX:
   - Integration with IBM products such as Watson Discovery, watsonx Assistant
   - Granite models and open-source models
   - Ease of using API/notebooks
   - Prompt tuning (scope carefully) a small model

watsonx.ai PoX flow – high-level view

- Client Discovery
  - Make sure that IBM and the client are on the same page
- **Planning the PoX.** Include elements to
  - Contain the scope including any overhead
  - Secure the needed resources
  - Maximize the chance of success
- Document of Understanding (DoU)
  - Ensure that the PoX is designed to achieve measurable metrics to meet the success criteria
  - Clearly states IBM and the client’s roles and responsibilities
- Execute the PoX
  - Work with the client to stay in scope
- After the PoX
  - Final presentation to sum up the PoX
  - Next step discussions

Client discovery – the most important step for watsonx.ai

- **What is the client's understanding of Generative AI?**

  - What is their experience with different models, model properties (size, configuration), prompt engineering?
  - Are the use cases for generative AI or traditional AI?
  - Are there any unrealistic expectations? (such as 100 % hallucination elimination)
  - What competitors are in place and are there any inaccurate or misleading messages that the client is buying into?
  - Sellers can start with this doc for discovery.

What gaps may need to be addressed before the PoX?

- IBM products (watsonx versus Watson, relationship to legacy products)
- Trade-offs between size, cost, performance, and creativity of models
- Strengths of generative AI versus traditional AI – neither replaces the other
- Benefits of LangChain, Retrieval Augmented Generation, and limitations
- Competitive analysis to understand what others offer and don't offer

Planning the PoX

The principle:

- Good planning is vital to the success of any PoX
- The scope must include capabilities that differentiate IBM from other vendors, or you have just spent weeks validating a lower-cost competitor’s solution

Critical questions:

- Which foundation model(s) available in watsonx.ai should be shortlisted for the PoX?
- Which other IBM products can best help reach the PoX goals while illustrating the capabilities of watsonx.ai?
- Where is the data required for the PoX and how would you set up access (once only)?
- What metrics need to be gathered along the way to prove the results of the PoX?
- What is required to measure the necessary metrics?

Planning the PoX – IBM and client roles/responsibilities

IBM

- Ensure an IBM Sales Cloud (ISC) record is in place to execute the PoX
- Ensure the use cases proposed meet the client's need and can be completed successfully within the time/resource limit
- Ensure the PoX can showcase IBM product's strengths and differentiation
- Ensure required resources are available
- Finalize the use cases with the client
- Agree on metrics and success criteria
- Select model(s) for the PoX based on use case requirements
- Agree to deliverables and next steps

Client

- Ensure that the proposed use cases satisfy the intended goals of the PoX
- Ensure that the required resources (software, data, environment, documents, data scientists, and other experts) are available for the duration of the PoX
- Finalize the use case together with IBM
- Finalize the metrics to measure, and sign off on the success criteria
- Agree to the deliverables and next steps

Preparing the Document of Understanding (DoU)

- An agreement between IBM and the client on the
  - Outlines the scope, timelines, success criteria, and roles and responsibilities for all parties.
    It should provide clarity to both parties on what happens upon successful completion of the PoX.

- A DoU should have most of the following general elements:
  - General background of the engagement
  - Activities and milestones to be performed
  - Expected deliverables;
  - Roles and responsibilities
  - Testing and metrics, and who is performing this
  - Success criteria
  - Architecture, including the environment, software, integrations, security and other dependencies
  - Assumptions and constraints

- Prevent scope creep

Document of understanding, along with examples

Links to samples in the speaker notes

Background

- ID Customer
- Summarize use case
- Reasons for doing the PoX
- What you want to accomplish
- What happens next

Activities

- Develop stack architecture
- Build and integrate with other applications/data
- Score performance
- Meet weekly
- Make final presentation

Deliverables

- Cloud Stack with generative AI addressing PoX requirements
- Final presentation PPT

Schedule

- High-level milestones and dates
- Begin work
- Kickoff meeting
- Checkpoints
- Testing Begin/End
- Complete deliverables
- Final presentation

Participants

- Who from IBM
- Who from client
- Roles/responsibilities

Testing Methods

- Who will test?
- In what environment?
- How will results be collected?
- How are results evaluated?
- What's out of bounds?

Scoring Methods

- What is a successful execution?
- What is a failure?
- What are you evaluating? (e.g., correct answers)
- What are you not evaluating? (e.g., UI)

Success Criteria

- What constitutes success overall?
- What KPIs, metrics determine success? (1 if possible, 3 is plenty)

PoX Architecture

- What environments?
- What software?
- What integrations?
- What data and where?
- What security?
- What dependencies on client tech?

Assumptions

- Scoping limits
- Make every contingency explicit
- What's in, what's out?
- What matters, what doesn't?
- Be explicit & verbose

“What’s on the menu?”

Things you can agree to without incurring too much additional risk and effort

watsonx. data

- Provides data management for any data that the client wants to use in a watsonx.ai PoX
- Depending on the timing of the PoX, watsonx.data provides integration to Milvus – a vector database that the client can use in a RAG use case

watsonx.governance

- Clients may sign onto the early access program for watsonx.governance, or simply use the GA version if the timing is right
- Watson.governance provides easy metrics calculation for watsonx.ai models, allowing a larger set of metrics to be included in evaluating models’ performance in a PoX
- Watsonx.governance provides many other capabilities to bolster the overall value-add to the client.

“What’s on the menu?”

Things you can agree to without incurring too much additional risk and effort

Watson Discovery

- Automate the discovery of information in the client’s repository. This is often useful in RAG use cases.
- Supports advanced NLP and NLU – providing NL interface together with watsonx.ai.
- Enterprise-grade, trusted, scalable and easy-to-use Natural Language AI.

watsonx Assistant

- Conversational AI platform. Watson Assistant supports easy chat integration with watsonx.ai.
- Empowers team to build AI-powered voice agents and chatbots to deliver automated self-service support.

watsonx Code Assist

- Leverages generative AI to accelerate code development
- Powered by the IBM Granite models
- Capabilities include
  - Code generation
  - Code matching
  - Code modernization

“What is a special order?”

Things you can agree to if you have access to skills, but which will add complexity, effort, time, and risk

Red flag if these are demanded for free on a deal under US$ 1 million

- Access to any client system, installation of any software in client environment
- Dependency on fee-based resources such as models hosted on other cloud providers
- NDA’s, No Charge PO’s, any legal document
- User authorizations, PII, privacy and security considerations

Executing the PoX and Post PoX

Executing the PoX

- Keep in mind that once the PoX kicks off, the clock is on
- Stay within the agreed scope, do NOT alter
  - Any exception requires mutual agreement between IBM and the client

Post PoX

- Summarize the findings of the PoX
  - Show how the PoX demonstrated the agreed-upon technical features through business scenarios or use cases
  - State objectives that have been successfully met
- Next steps
  - At the successful completion of a PoX, ask for the next steps in the sales cycle (should have been documented in the DoU)

Additional resources

As you prepare to execute your watsonx.ai PoX, make a note of these resources available to help you.

- You and your manager should review the guidance for making the most out of the PoX.

- Various demos of watsonx.ai
  - watsonx.ai solution gallery
  - AI for Business demo
  - IBM watsonx Demo Hub

- Product documentation for watsonx and watsonx.ai

- Watsonx.ai Sales Kit on Seismic

- Once you exhaust these, seek assistance from the watsonx.ai community and the resources they contribute to:
  - Support Slack channel: #watsonx-ai-enablement
  - The Generative AI Series on Events and Classes
  - TechZone support: https://techzone.ibm.com/help
  - TechZone community: #itz-techzone-community