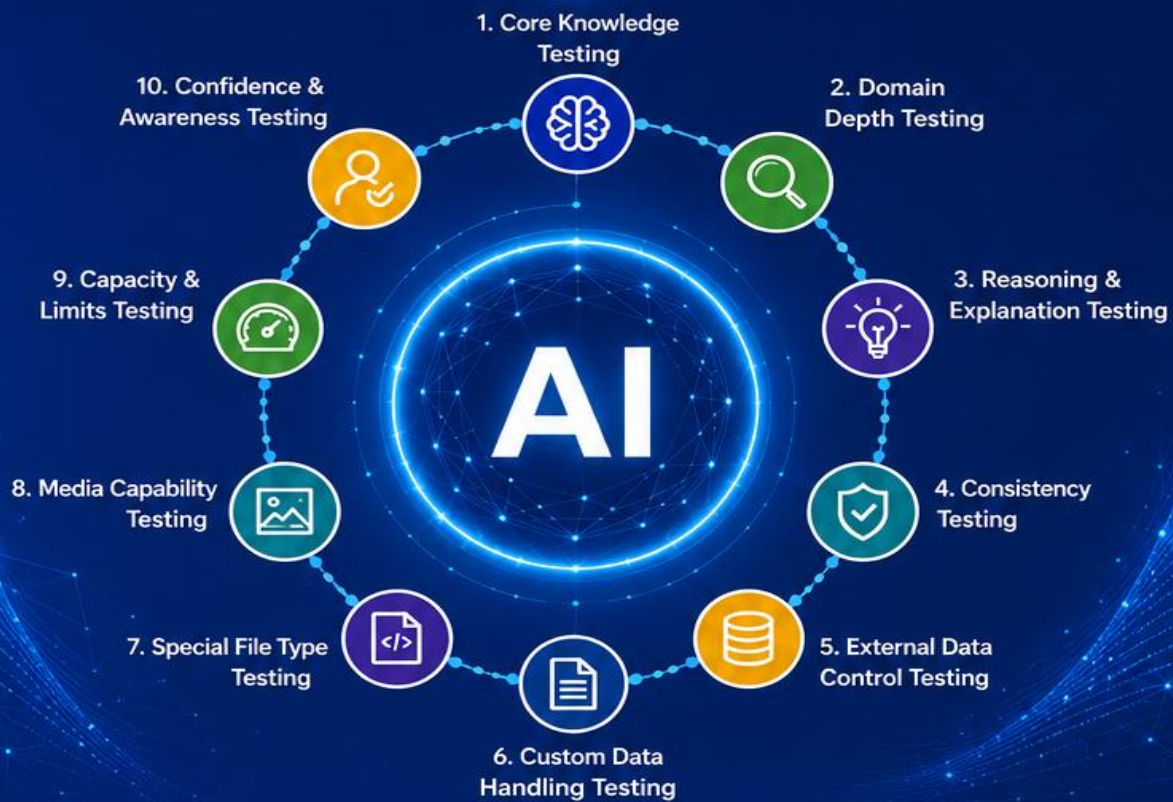


Before You Trust AI:

A Practical Framework to Test What It **Really** Knows and Can Do



Know its strengths



Understand its limits



Test with real scenarios



Make informed decisions



Unlock real value



By Assem Hijazi



<https://www.linkedin.com/in/assemhijazi/>

Before You Trust AI: A Practical Framework to Test What It Really Knows and Can Do

Contents

- Introduction 7
 - What This Article Will Do 7
 - The Key Idea 7
- Executive Summary 8
 - AI Evaluation Framework 8
 - Key Executive Insights 9
 - Business Value 9
 - Final Message..... 9
- Executive summary diagram 10
- The AI Testing Framework (Overview Table)..... 11
 - AI Assistant Testing Framework 11
 - Simple Way to Understand the Framework..... 11
 - Key Insight..... 12
- 1. Core Knowledge Testing (What the AI Really Knows) 13
 - What Is Core Knowledge Testing? 13
 - How to Test It (Simple Method)..... 13
 - What You Should Look For..... 13
 - What This Test Reveals 13
 - Important Insight 14
 - Simple Example 14
 - Common Mistake to Avoid 14
 - Key Takeaway 14
- 2. Domain Depth Testing (How Deep the AI Really Understands a Topic) 15
 - What Is Domain Depth Testing? 15
 - How to Do It (Step-by-Step) 15
 - Step 1: Start with a General Question..... 15
 - What You Are Testing..... 15
 - What This Test Reveals 16

Important Insight	16
What Happens in Weak Areas	16
Advanced Variation (Very Powerful)	16
This tests:	16
Key Takeaway	17
3. Reasoning & Explanation Testing (Does the AI Really Understand?)	18
What Is Reasoning Testing?	18
How to Do It (Simple Methods).....	18
Example Prompts	18
What You Are Testing	18
What Strong Reasoning Looks Like	18
What Weak Reasoning Looks Like	19
Advanced Testing (Very Powerful)	19
This tests:	19
Important Insight	19
Key Takeaway	19
4. Consistency Testing (Is the AI Stable or Contradicting Itself?)	20
What Is Consistency Testing?.....	20
How to Do It (Simple Methods).....	20
Example Prompts	20
What You Are Testing	20
What Strong Consistency Looks Like	20
What Weak Consistency Looks Like	21
Advanced Testing (Very Useful)	21
This helps detect:	21
Important Insight	21
Why This Matters	21
Key Takeaway	22
5. External Data Control Testing (Can You Guide Where AI Gets Its Answers?)	23
What Is External Data Control Testing?	23
Two Levels of Source Control	23
Level 1: Guided Control (Prompt-Based)	23
Level 2: Strict Control (System-Based).....	24

What This Enables	24
What You Are Testing.....	24
What Strong Performance Looks Like	24
What Weak Performance Looks Like.....	24
Important Insight	25
Why This Changes How You Use AI.....	25
Key Takeaway	25
Example of Level 2: Strict Source Control Using a Website List	26
Example Excel Structure.....	26
What Happens in This Case	26
Example Prompt	26
What You Achieve	26
Important Clarification	27
Simple Way to Explain It in Your Article	27
Powerful Line You Can Use	27
This example is excellent because:	27
6. Custom Data Handling Testing (How Well AI Works with Your Files)	28
What Is Custom Data Handling Testing?	28
How to Test It (Simple Methods)	28
1. Understanding Test	28
2. Analysis Test.....	29
3. Data Extraction Test	29
4. Cross-File Testing (Advanced)	29
What You Are Testing.....	29
What Strong Performance Looks Like	29
What Weak Performance Looks Like.....	30
Important Insight	30
Advanced Testing Idea (Very Useful)	30
Key Takeaway	30
7. Special File Type Testing (Advanced and Non-Standard Data).....	31
What Is Special File Type Testing?	31
How to Test It (Simple Approach).....	31
What You Are Testing.....	31

What You Will Often Discover (Important Reality)	32
Why This Happens	32
Strong vs Weak Capability	32
Strong Capability	32
Weak Capability	32
Advanced Insight	32
Practical Workaround (Very Useful)	32
Key Takeaway	33
8. Media Capability Testing (Images, Audio, and Video)	34
What Is Media Capability Testing?	34
How to Test It (Simple Methods)	34
1. Image Understanding Test	34
2. Image Analysis (Advanced)	35
3. Audio Testing	35
4. Video Testing (Advanced / Limited in Many Cases)	35
What You Are Testing	36
What Strong Performance Looks Like	36
What Weak Performance Looks Like	36
Important Insight	36
Practical Tip.....	36
Key Takeaway	36
9. Capacity & Limits Testing (Discovering the Real Boundaries of AI).....	37
What Is Capacity & Limits Testing?	37
Step 1: Ask the AI About Its Limits (Very Important First Step).....	37
Step 2: File Count and Size Testing	37
Step 3: Multi-File Combination Testing	38
Step 4: Token and Context Awareness (Advanced Testing)	38
Step 5: Media and File Size Limits	38
What You Are Really Testing	39
Key Insight.....	39
What Strong Capability Looks Like	39
What Weak Capability Looks Like	39
Practical Tip.....	39

Key Takeaway	40
10. Confidence & Awareness Testing (Does the AI Know Its Limits?)	41
What Is Confidence & Awareness Testing?	41
How to Test It (Simple Methods)	41
1. Ask for Confidence Level	41
2. Ask for Limitations	41
3. Challenge the Answer	42
4. Ask About Missing Information	42
What You Are Testing	42
What Strong Performance Looks Like	42
What Weak Performance Looks Like	42
Important Insight	43

Introduction

Today, AI assistants are everywhere. People use them to: Answer questions, Analyze documents, Support business decisions, Even generate ideas and strategies

But there is one problem: Most people use AI... without ever testing it. They assume: The AI understands everything, It always gives accurate answers, It can handle any type of data, But in reality, every AI assistant has: Strengths, Limitations, And very specific capabilities

The challenge is: These differences are not always visible in the answer. Two AI assistants may give similar responses... but their **real capabilities can be completely different**

If you don't test your AI assistant: You may trust weak answers, You may use it in the wrong scenarios, You may miss its real potential

What This Article Will Do

Instead of explaining AI from a theoretical perspective... This article gives you a **practical framework**, A simple way to test:

- What the AI really knows
- How it handles different types of data
- How deep its understanding goes
- And where its limits are

The Key Idea

Before you trust AI... you need to understand it, Not by reading about it... But by testing it the right way

Executive Summary

AI assistants are widely used across business and personal applications, yet most users rely on them, without clearly understanding their capabilities, limitations, or reliability.

This document introduces a **practical evaluation framework** to help users test AI assistants in a structured way—ensuring better decisions, higher trust, and more effective usage.

AI Evaluation Framework

#	Testing Area	Objective	What It Reveals
1	Core Knowledge	Test what AI knows without external help	Strength of the base model
2	Domain Depth	Test how deep AI understands a topic	Where knowledge is strong vs weak
3	Reasoning & Explanation	Test how AI thinks and explains	Real understanding vs surface answers
4	Consistency	Test stability across different prompts	Reliability of responses
5	External Data Control	Test ability to guide data sources	Level of trust and control
6	Custom Data Handling	Test handling of user files	Practical usability with business data
7	Special File Types	Test advanced formats (CAD, medical, etc.)	Capability limits in specialized domains
8	Media Capabilities	Test images, audio, video understanding	Multimodal intelligence level
9	Capacity & Limits	Test data size, files, and context limits	Operational boundaries
10	Confidence & Awareness	Test AI’s ability to recognize uncertainty	Risk awareness and reliability

Key Executive Insights

Area	Insight
Reliability	A single correct answer does not guarantee reliability—consistency is critical
Data Usage	More data does not always improve results—understanding matters more
Hidden Limits	AI systems operate within limits (data, context, files) that directly affect outcomes
Control	Guiding or controlling data sources significantly improves trust and accuracy
Risk	Overconfidence without awareness is one of the biggest risks in AI usage

Business Value

Benefit	Impact
Better Decision-Making	Reduced risk of relying on weak or incorrect outputs
Improved AI Adoption	Clear understanding of where AI adds value
Higher Trust	Controlled and validated AI usage
Operational Efficiency	Better use of AI within its real capabilities
Strategic Use	Moving from casual use → structured AI integration

Final Message

AI should not be trusted by default—it must be tested, understood, and controlled

Before You Trust AI

A Practical Framework to Test What It Really Knows and Can Do

Most People Use AI Without Testing It

- Assume It Knows Everything
- Trust Answers Blindly
- Ignore Hidden Limits

Result: Risky Decisions & Unreliable Outcomes

AI Testing Framework

10 Key Areas to Evaluate AI



Core Knowledge



Domain Depth



Reasoning



Consistency



Data Control



Custom Data



Special Files



Media Skills



Capacity Limits



Awareness

How to Use It

1. Test What It Knows ➤ **2. Test How It Thinks** ➤ **3. Test What It Can Handle** ➤ **4. Test Its Limits**

Critical Insights

- AI Isn't Always Right
- More Data $\neq$ Better Results
- Limits Affect Outcomes
- Control Builds Trust

Business Benefits

- Better Decisions
- Higher Trust
- Reduced Risk
- Smarter AI Use



Don't Just Use AI... Test It, Understand It, and Control It.

The AI Testing Framework

To properly evaluate any AI assistant, we can break the process into a set of **practical testing areas**. Each area focuses on a different capability—together, they give a complete picture of what the AI can really do.

AI Assistant Testing Framework

#	Testing Area	What You Test	Why It Matters (Quick Description)
1	Core Knowledge Testing	What the AI knows without external help	Helps you understand the strength of the base model before adding data
2	Domain Depth Testing	How deep the AI understands a specific domain	Shows where the knowledge is strong and where it becomes weak
3	Reasoning & Explanation Testing	How well the AI explains and reasons	Distinguishes real understanding from surface-level answers
4	Consistency Testing	Stability of answers across different prompts	Detects contradictions or unstable knowledge
5	External Data Control Testing	Ability to guide AI to specific sources	Builds trust by controlling where answers come from
6	Custom Data Handling Testing	Handling of files (PDF, Word, Excel, etc.)	Validates how well AI works with user-provided data
7	Special File Type Testing	Handling of advanced formats (CAD, medical, etc.)	Identifies limitations in non-standard data processing
8	Media Capability Testing	Understanding images, audio, and video	Tests multimodal capabilities of the AI assistant
9	Capacity & Limits Testing	Data size, context limits, and usage constraints	Helps you understand practical limits per chat or subscription
10	Confidence & Awareness Testing	How AI handles uncertainty	Shows whether AI recognizes its own limitations

Simple Way to Understand the Framework

- First → Test what the AI knows
- Then → Test how it thinks
- Then → Test what it can handle

- Finally → Test its limits
-

Key Insight

| Using AI without testing these areas is like using a tool without knowing what it can or cannot do.

1. Core Knowledge Testing (What the AI Really Knows)

Before testing anything else, there is one critical step: **Understand what the AI already knows on its own.** This is the foundation of everything.

What Is Core Knowledge Testing?

Core knowledge testing means: Asking the AI to answer using only its internal trained knowledge, Without relying on external data, documents, or browsing

Most users skip this step and immediately: Upload files, Add data and Use advanced tools, but this creates a problem: You don't know if the AI truly understands the topic... or if it is just relying on external support

How to Test It (Simple Method)

Start with a clear and controlled prompt.

Example Prompts

- “Answer this question using only your internal knowledge. Do not use external data or additional sources.”
 - “Explain this topic based only on your trained knowledge.”
 - “Give me your answer without relying on documents or browsing.”
-

What You Should Look For

When you get the answer, evaluate:

- **Clarity** → Is the explanation structured and easy to follow?
 - **Depth** → Does it go beyond basic definitions?
 - **Accuracy** → Are the key points correct?
 - **Confidence** → Does it sound unsure or inconsistent?
-

What This Test Reveals

This simple step helps you understand:

- The **strength of the base model**

- The **domains where it performs well**
 - The **areas where it may struggle**
-

Important Insight

| AI does not become smarter by adding data first, You need to understand its starting point

Simple Example

Ask: “Explain how blood circulates in the human heart.”

Then observe:

- Does it explain clearly step by step?
 - Does it use correct terminology?
 - Does it go deep or stay basic?
-

Common Mistake to Avoid

Many users assume: “If AI gives an answer, it must understand the topic”. But in reality: It may only have **partial understanding**

Key Takeaway

| **Always test the core knowledge first**, It defines how reliable everything else will be

2. Domain Depth Testing (How Deep the AI Really Understands a Topic)

After testing the core knowledge, the next step is: **Testing how deep that knowledge goes**, Because knowing something at a surface level is very different from truly understanding it.

What Is Domain Depth Testing?

This test checks: How far the AI can go from general knowledge → to highly specific understanding

How to Do It (Step-by-Step)

You start broad... then go deeper step by step.

Step 1: Start with a General Question

Example: “Explain healthcare systems.”

Step 2: Narrow the Domain

“Explain heart diseases.”

Step 3: Go Deeper into a Subdomain

“Explain how blood circulates in the human heart.”

Step 4: Go Very Specific

“Explain how oxygenated and deoxygenated blood move through the left and right sides of the heart, step by step.”

What You Are Testing

At each level, observe:

- Does the explanation remain clear?
- Does the detail improve or become weaker?
- Does the AI stay accurate?

What This Test Reveals

This helps you identify:

- Where the AI is **strong**
- Where it becomes **generic**
- Where it starts to **lose accuracy or clarity**

Important Insight

Strong AI is not defined by what it knows broadly...
but by how deep it can go without losing quality

What Happens in Weak Areas

When the AI reaches the limit of its knowledge:

- Answers become more general
- Explanations lose structure
- Details may become incorrect or vague

Advanced Variation (Very Powerful)

You can push even further by asking:

- “Explain this at an expert level”
- “What are the edge cases or rare scenarios?”
- “What are common misconceptions in this topic?”

This tests:

- Depth of training
 - Exposure to complex knowledge
 - Ability to go beyond basics
-

Key Takeaway

Depth matters more than surface knowledge, This test shows where AI truly understands—and where it only approximates

3. Reasoning & Explanation Testing (Does the AI Really Understand?)

After testing knowledge and depth, the next step is: **Testing how the AI thinks—not just what it knows**

Because knowing information is one thing... **Explaining and reasoning with it is something else**

What Is Reasoning Testing?

This test checks: Whether the AI can **analyze, explain, and connect ideas logically**, Not just repeat information.

How to Do It (Simple Methods)

Instead of asking for definitions, ask for reasoning.

Example Prompts

- “Explain *why* this happens, not just what it is.”
 - “Compare these two approaches and explain which one is better and why.”
 - “Walk me through this step by step.”
 - “What would happen if we change this condition?”
-

What You Are Testing

You are looking for:

- **Logical flow** → Does the answer follow a clear reasoning path?
 - **Cause and effect** → Does it explain why things happen?
 - **Connections** → Does it link ideas together?
 - **Clarity** → Is the explanation understandable or confusing?
-

What Strong Reasoning Looks Like

- Step-by-step explanation
- Clear cause → effect relationships
- Structured thinking

- Ability to compare and evaluate
-

What Weak Reasoning Looks Like

- Repeating definitions
 - Jumping between ideas without logic
 - Missing “why” explanations
 - Giving general or vague answers
-

Advanced Testing (Very Powerful)

Push the AI further by asking:

- “Challenge your own answer—what could be wrong?”
- “Give an alternative explanation.”
- “Explain this from another perspective.”

This tests:

- Depth of reasoning
 - Flexibility of thinking
 - Ability to self-evaluate
-

Important Insight

AI can sound confident even when reasoning is weak, That’s why this test is critical: It separates **real understanding** from **surface-level answers**

Key Takeaway

Don’t just test what AI knows—test how it thinks, Strong reasoning is the real indicator of intelligence

4. Consistency Testing (Is the AI Stable or Contradicting Itself?)

After testing knowledge and reasoning, the next step is: **Testing whether the AI is consistent in its answers**, because a strong AI should not just give a good answer once... It should give **reliable and stable answers across different prompts**

What Is Consistency Testing?

This test checks: Whether the AI gives **aligned answers** when the same idea is asked in different ways

How to Do It (Simple Methods)

Ask the same concept using different styles:

Example Prompts

1. Direct question

“What is the function of the heart?”

2. Rephrased question

“What does the heart do in the human body?”

3. Scenario-based

“Explain the role of the heart in blood circulation.”

4. Detailed request

“Describe step by step how the heart supports the body.”

What You Are Testing

Look for:

- **Alignment** → Are the answers consistent in meaning?
 - **No contradictions** → Does it change facts or explanations?
 - **Stable quality** → Does quality stay the same or fluctuate?
-

What Strong Consistency Looks Like

- Same core explanation across prompts

- Details may vary, but the meaning stays correct
 - No contradictions
-

What Weak Consistency Looks Like

- Conflicting information
 - Changing explanations
 - Different conclusions for the same concept
-

Advanced Testing (Very Useful)

Try:

- Asking the same question in different sessions
 - Changing wording significantly
 - Asking follow-up questions that challenge previous answers
-

This helps detect:

- Instability in knowledge
 - Weak understanding
 - Hidden inconsistencies
-

Important Insight

A single good answer does not mean the AI is reliable, Reliability comes from: **Consistency over multiple interactions**

Why This Matters

In real use:

- Business decisions
- Technical analysis
- Medical or financial advice

Inconsistency can lead to serious issues

Key Takeaway

| Trust AI only when it is consistent—not just when it sounds correct once

5. External Data Control Testing (Can You Guide Where AI Gets Its Answers?)

After testing knowledge, reasoning, and consistency, the next step is: **Testing whether you can control where the AI gets its information from**

This is a key step for:

- Building trust
 - Improving accuracy
 - Using AI in business or critical scenarios
-

What Is External Data Control Testing?

This test checks: Whether the AI can follow instructions to use **specific and trusted sources of information**, Instead of giving general answers.

By default, AI answers are based on:

- Core knowledge
- General patterns
- Mixed sources

But in many cases, you need: **Controlled, reliable, and trusted sources**

Two Levels of Source Control

This is where your approach becomes very powerful.

Level 1: Guided Control (Prompt-Based)

At this level, you guide the AI using prompts.

Example Prompts

- “Answer this question using only trusted healthcare sources like WHO or Mayo Clinic.”
- “Base your answer on academic and research-level information.”
- “Align your response with industry best practices and official guidelines.”

What Happens Here

- The AI **does not literally browse only those websites**
- It **adjusts its answer style and content** to match those types of sources

You are influencing the answer—not strictly controlling it

Level 2: Strict Control (System-Based)

At this level, real control is applied through systems such as:

- RAG (Retrieval-Augmented Generation)
 - Connected databases
 - Enterprise knowledge platforms
-

What This Enables

- Limiting the AI to **specific websites or datasets**
- Using **approved company or industry sources**
- Ensuring **high accuracy and compliance**

This is **true source control**

What You Are Testing

- Can the AI follow source guidance?
 - Does the answer become more structured and reliable?
 - Does the tone shift to match trusted sources?
-

What Strong Performance Looks Like

- More structured and professional answers
 - Alignment with trusted knowledge styles
 - Reduced randomness or unsupported claims
-

What Weak Performance Looks Like

- Ignoring source instructions
 - Mixing unrelated or informal content
 - No noticeable improvement in answer quality
-

Important Insight

“You can guide AI with trusted sources—but true control comes from connecting it to those sources directly.”

Why This Changes How You Use AI

Instead of accepting any answer... You start **shaping the quality and reliability of the response**

This is especially important for:

- Business decisions
 - Research
 - Healthcare and regulated industries
-

Key Takeaway

AI becomes more reliable when you guide it—and even more powerful when you control its data sources

Example of Level 2: Strict Source Control Using a Website List

In advanced scenarios, instead of just guiding the AI with prompts... You can **control exactly where the AI gets its information from**

Scenario You provide the AI assistant with: An **Excel file** that contains a list of trusted websites

Example Excel Structure

#	Website Name	URL	Category
1	WHO	https://www.who.int	Healthcare
2	Mayo Clinic	https://www.mayoclinic.org	Healthcare
3	CDC	https://www.cdc.gov	Healthcare
4	NHS	https://www.nhs.uk	Healthcare

What Happens in This Case

When integrated properly (via RAG or connected system):

- The AI **reads the Excel file**
 - It understands that these are the **allowed sources**
 - It retrieves information only from these websites
 - It builds answers based on **controlled and trusted data**
-

Example Prompt

“Answer the question using only the websites listed in the uploaded Excel file.”

What You Achieve

- High trust in the answer
 - Controlled and validated sources
 - Reduced risk of misinformation
-

Important Clarification

This works as **true control** only when:

- The system supports **RAG or data connections**
- The AI can actually **access and retrieve from those sources**

Otherwise: The Excel file becomes **guidance, not strict enforcement**

Simple Way to Explain It in Your Article

- Prompt only → “Use trusted sources” (Guidance)
 - Excel + RAG → “Use these exact sources” (Control)
-

Powerful Line You Can Use

“When you move from telling AI what to use... to giving it exactly where to look—you move from influence to control.”

This example is excellent because:

- It connects theory → real implementation
 - It fits enterprise use cases perfectly
 - It shows how AI can be governed, not just used
-

6. Custom Data Handling Testing (How Well AI Works with Your Files)

After testing how AI uses external sources, the next step is: **Testing how well the AI can work with your own data**, because in real use, AI is not just answering questions... It is expected to **read, understand, and analyze your files**

What Is Custom Data Handling Testing?

This test checks: How effectively the AI can process and use **user-provided files**

Such as:

- PDF documents
- Word files (DOCX)
- Excel sheets (XLSX, CSV)
- PowerPoint presentations (PPTX)

In many real-world scenarios:

- Business reports are in PDF
- Data is in Excel
- Presentations are in PowerPoint

If the AI cannot handle these properly, its value becomes limited

How to Test It (Simple Methods)

Upload different types of files and ask the AI to perform tasks.

Example Tests

1. Understanding Test

“Summarize this document.”

Tests:

- Reading ability
 - Key point extraction
-

2. Analysis Test

“What are the key insights from this report?”

Tests:

- Interpretation
 - Analytical capability
-

3. Data Extraction Test

“Extract the main numbers and present them in a table.”

Tests:

- Accuracy
 - Structure handling
-

4. Cross-File Testing (Advanced)

Upload multiple files and ask:

“Compare the findings in these two documents.”

Tests:

- Multi-source understanding
 - Context linking
-

What You Are Testing

- Can the AI read the file correctly?
 - Does it understand the structure?
 - Can it extract meaningful insights?
 - Can it connect information across files?
-

What Strong Performance Looks Like

- Clear and structured summaries
- Accurate extraction of key points
- Ability to analyze and compare

- Understanding of tables, charts, and text
-

What Weak Performance Looks Like

- Missing important information
 - Misinterpreting data
 - Giving generic summaries
 - Ignoring structure (tables, sections)
-

Important Insight

Not all files are equal—how the data is structured affects how well AI can understand it

Examples

- Clean Excel sheet → easy to analyze
 - Scanned PDF → harder to interpret
 - Complex formatting → may reduce accuracy
-

Advanced Testing Idea (Very Useful)

Try mixing:

- Clean structured files
- Messy or unstructured files

This shows the **real capability limits**

Key Takeaway

AI is only as good as its ability to understand your data, Testing file handling is essential for real-world use

7. Special File Type Testing (Advanced and Non-Standard Data)

After testing common files like PDF and Excel, the next step is: **Testing how the AI handles specialized and non-standard file types**

This is where many AI assistants start to show clear limitations.

What Is Special File Type Testing?

This test checks: Whether the AI can understand and work with **advanced or industry-specific file formats**

Such as:

- CAD drawings (engineering, construction)
- Medical imaging files (MRI, CT scans)
- Scientific datasets
- Technical design files

In many industries:

- Engineering uses CAD files
- Healthcare uses imaging and diagnostic data
- Research uses specialized formats

If the AI cannot handle these formats, its role becomes limited in those domains

How to Test It (Simple Approach)

Try uploading or describing these file types and ask:

Example Prompts

- “Can you analyze this CAD drawing and explain the structure?”
 - “What insights can you extract from this medical imaging file?”
 - “Can you interpret this technical design file?”
-

What You Are Testing

- Does the AI recognize the file type?
- Can it interpret the structure or content?

- Does it provide meaningful insights or just general responses?
-

What You Will Often Discover (Important Reality)

In many cases: The AI will not fully process these files directly

Instead:

- It may give general explanations
 - It may require converted or simplified formats
 - It may depend on integrations with specialized tools
-

Why This Happens

Because: These file types require **domain-specific processing tools**, not just language understanding

Strong vs Weak Capability

Strong Capability

- Can interpret structured technical data
 - Provides meaningful insights
 - Understands domain context
-

Weak Capability

- Gives generic explanations
 - Cannot read or process the file
 - Requires manual conversion
-

Advanced Insight

AI assistants are strongest in language and structured data, but may require additional systems for specialized formats

Practical Workaround (Very Useful)

Instead of using raw files:

- Convert CAD → images or simplified formats
- Convert medical data → reports or summaries
- Extract key data before giving it to AI

This often improves results significantly

Key Takeaway

Not all data types are equally supported by AI, Specialized formats often require additional tools or preparation

8. Media Capability Testing (Images, Audio, and Video)

After testing documents and special file types, the next step is: **Testing how the AI handles media content**, because modern AI assistants are no longer limited to text... They are expected to **see, hear, and interpret media**

What Is Media Capability Testing?

This test checks:

How well the AI can understand and work with:

- Images
- Audio (voice recordings)
- Video content

In real-world scenarios:

- Users share screenshots instead of text
- Businesses use images and videos for analysis
- Communication is increasingly visual and audio-based

AI must go beyond text to be truly useful

How to Test It (Simple Methods)

1. Image Understanding Test

Upload an image and ask:

- “What do you see in this image?”
- “Explain this chart or diagram.”
- “What insights can you extract from this screenshot?”

Tests:

- Object recognition
 - Context understanding
 - Ability to interpret visual data
-

2. Image Analysis (Advanced)

- “What problems do you see in this design?”
- “Analyze this UI and suggest improvements.”

Tests:

- Analytical thinking on visual input
 - Practical reasoning
-

3. Audio Testing

If supported, provide:

- Voice input
- Audio recordings

Ask:

- “Summarize this audio.”
- “What are the key points discussed?”

Tests:

- Speech understanding
 - Content extraction
-

4. Video Testing (Advanced / Limited in Many Cases)

Ask:

- “Can you analyze this video?”
- “Summarize the key moments in this video.”

In many cases:

- AI may not directly process full video
 - It may rely on:
 - Transcripts
 - Extracted frames
-

What You Are Testing

- Can the AI understand visual and audio input?
 - Does it extract meaningful insights?
 - Can it go beyond description into analysis?
-

What Strong Performance Looks Like

- Clear and accurate descriptions
 - Ability to interpret charts, UI, and visuals
 - Extracting insights—not just describing content
-

What Weak Performance Looks Like

- Basic or obvious descriptions only
 - Missing important details
 - No deeper analysis
-

Important Insight

Seeing is not understanding. Some AI systems can describe images... But not all can **analyze or interpret them deeply**

Practical Tip

For better results:

- Provide clear images
 - Add context with your prompt
 - Ask specific questions (not just “what is this?”)
-

Key Takeaway

Multimodal AI adds powerful capabilities—but it must be tested carefully, True value comes from understanding, not just recognition

9. Capacity & Limits Testing (Discovering the Real Boundaries of AI)

After testing knowledge, data handling, and media capabilities, the next step is: **Understanding the real limits of the AI assistant**, but instead of guessing those limits... You should **ask the AI directly and then validate it**

What Is Capacity & Limits Testing?

This test focuses on:

Discovering how much data the AI can handle

Understanding system constraints based on usage and subscription

Step 1: Ask the AI About Its Limits (Very Important First Step)

Before testing anything, start with direct questions.

Example Questions

- “How many pages of documents can you handle in one chat?”
- “What is your maximum context size for this session?”
- “How does this differ based on subscription or model?”

The AI should provide:

- Estimated limits
- Differences between plans or models
- General capacity guidelines

Important Reality

Many AI assistants may not give exact or fully accurate answers, But: This still gives you a **baseline understanding**

Step 2: File Count and Size Testing

Now go deeper:

Ask About File Limits

- “How many files can you process in a single session?”
- “What is the maximum size per file?”

Test It Practically

- Upload multiple files
 - Increase size gradually
 - Observe when performance drops
-

Step 3: Multi-File Combination Testing

Once limits are understood: Test real usage

Example Prompt

“Combine insights from all these files and give a unified summary.”

This tests:

- Cross-file understanding
 - Context management
 - Stability under load
-

Step 4: Token and Context Awareness (Advanced Testing)

This is one of your strongest ideas.

Ask During the Conversation

- “How much of the context have we used so far?”
- “Are we approaching any limits in this chat?”

Reality Check

Most AI assistants:

- Do not show exact token usage
- May give approximate or general answers

But: This test checks **awareness of limits**, not just the limits themselves

Step 5: Media and File Size Limits

You can also test:

Example Questions

- “What is the maximum size of an image you can analyze?”

- “Can you process large video files? If yes, what is the limit?”
- “What types of media files are supported?”

This helps you understand:

- Practical boundaries
 - Supported formats
 - Performance expectations
-

What You Are Really Testing

This section is not just about limits... It is about **transparency and control**

- Does the AI clearly explain its limits?
 - Can you plan your usage based on those limits?
 - Can you avoid performance issues?
-

Key Insight

| The limits of AI are often hidden—but they directly affect the quality of results

What Strong Capability Looks Like

- Clear explanation of limits
 - Stable performance under load
 - Ability to handle multiple files and large inputs
-

What Weak Capability Looks Like

- Vague or unclear answers about limits
 - Sudden performance drops
 - Loss of context without warning
-

Practical Tip

Always ask about limits first... then test them

This avoids:

- Overloading the system
 - Losing important data
 - Getting incomplete answers
-

Key Takeaway

Understanding AI limits is not optional—it is essential for reliable use

The best users don't just use AI... they work within its boundaries

10. Confidence & Awareness Testing (Does the AI Know Its Limits?)

After testing knowledge, reasoning, data handling, and system limits, the final step is: **Testing whether the AI understands its own limitations.** because one of the biggest risks when using AI is not incorrect answers... It is **confident answers that may not be fully correct**

What Is Confidence & Awareness Testing?

This test checks:

Whether the AI can recognize:

- Uncertainty
- Gaps in knowledge
- Limits in its capabilities

In real-world usage:

- Users often trust AI because it sounds confident
- But confidence does not always mean accuracy

This is where many mistakes happen

How to Test It (Simple Methods)

1. Ask for Confidence Level

“How confident are you in this answer and why?”

You are testing:

- Self-evaluation
 - Awareness of uncertainty
-

2. Ask for Limitations

“What are the possible limitations of your answer?”

Strong AI should:

- Highlight assumptions
 - Mention possible gaps
-

3. Challenge the Answer

“What could be wrong in your answer?”

“Give an alternative explanation.”

This tests:

- Critical thinking
 - Ability to question itself
-

4. Ask About Missing Information

“What additional data would improve your answer?”

This reveals:

- Awareness of what is not known
 - Ability to guide the user
-

What You Are Testing

- Does the AI admit uncertainty?
 - Does it explain limitations clearly?
 - Can it evaluate its own response?
-

What Strong Performance Looks Like

- Clear acknowledgment of uncertainty
 - Honest explanation of limitations
 - Balanced and realistic answers
-

What Weak Performance Looks Like

- Overconfidence in all answers
 - No mention of limitations
 - Avoiding uncertainty
-

Important Insight

The smartest AI is not the one that answers everything...but the one that knows when it might be wrong