



# General-purpose AI as an assistant in the bioanalytical lab

Annika Carlsson, EBF Open Symposium 2025

# Comparison of General-purpose AI (GP) & Specialized-purpose AI (SP)

## ☑ Examples of systems

GP – ChatGPT, Claude, Gemini, Mistral, Perplexity, MS Co-pilot

SP – Cell image software, AI-enabled freezers, AI-assisted curve fit programs

## ☑ Scope

GP – Broad, cross-domain  
SP – Narrow, single domain

## ☑ Human interaction

GP – Conversational & flexible  
SP – Usually less interactive

## ☑ Training

GP – On diverse data  
SP – On domain-specific data

## ☑ Adaptability & control

GP – High adaptability, learns from context, higher risk of hallucination, need supervisory control  
SP – Low adaptability, fixed to training tasks, lower risk of hallucination, rely on deterministic control

## ☑ Performance

GP – Moderate across many tasks  
SP – Excellent on specialized tasks

## ☑ Maintenance

GP – One model serves many use cases  
SP – Specific model for use case

## ☑ Examples of tasks

GP – Report generation, assess data trends, explain deviations  
SP – image segmentation, sample integrity monitoring, outlier handling



# General-purpose AI in the enterprise landscape

System category	Definition	Governance	Control level	Examples
In-house / Private AI	AI models hosted entirely within organizations infrastructure	All data, model weights and logs remain internal	Full	Company-trained LLMS, on-premises clones of GPT's
Enterprise-embedded AI systems	AI assistants integrated within enterprise platforms	Inherit corporate identity, permissions and tenant security	High via tenant and permissions	Microsoft 365 Copilot, GitHub copilot
Managed external AI services	Vendor-hosted AI platforms designed for enterprise security	By contract and compliance controls	Partial, external hosting with contract	ChatGTP enterprise, Claude enterprise, Gemini enterprise
Public general purpose AI systems	Moderate across many tasks	Shared cloud environment, require user caution and data anonymization	Minimal	ChatGPT Free/plus, Gemini free, Perplexity, Poe



**Case study**  
**MS Co-pilot as an assistant for  
data analysis and data  
interpretation**



## **Prompt:**

Using the Excel file (specified), write a lab journal report for screening of samples following the template presented in word (specified). Please show me the text for revision.

# AI output



## Labjournal

Title: Analysis of [A] Total [B] and [C] Levels

Document No.: Lj2025-001

Page: 1(1)

Author: [Your Name]

Signature: [Your Signature]

Dep.: [Your Department]

Date: [Date]

## Purpose & Scope

The purpose of this study is to analyze the levels of [A] Total [B] and [C] in various samples using [X] platform. The scope includes the preparation of reagents, execution of the assay, and analysis of the results to determine the concentration of these biomarkers in the samples.

## Material & Equipment

- [NN1] Plate [Lot No]
- [NN2] [Lot No]
- [C1] [NN3] [Lot No]
- [NN5] [Lot No]
- [C2] Antibody [Lot No]
- Aprotinin [Lot No]
- [B1] Antibody [Lot No]
- [NN6] Buffer [Lot No]
- [B2] Antibody [Lot No]
- [NN7] Solution [Lot No]
- [A1] Antibody [Lot No]
- Calibrator [Lot No]
- [A2] Antibody [Lot No]
- Wash Buffer [Lot No]
- Various pipettes (M-1145, M-1133, M-1136, M-39, M-1142)
- Washer (M-1125)
- Plate Reader (M-1179)

# AI output



## Description

### 1. Preparation of Reagents:

- Dilute [NN] to 1X using diH2O.
- Prepare antibody mixtures for coating by [NN] and stop solutions.
- Prepare metabolic assay working solution and calibrators.

### 2. Assay Procedure:

- Add 200  $\mu\text{L}$  of [NN] to 300  $\mu\text{L}$  of [NN] solutions.
- Vortex and incubate at room temperature for 30 minutes.
- Add 50  $\mu\text{L}$  of [NN] to each well and seal the plate. Incubate at room temperature with shaking for 1 hour.
- Wash the plate 3 times with 300  $\mu\text{L}$ /well of wash buffer.
- Add 50  $\mu\text{L}$  of calibrator, control, or sample to the wells and incubate at room temperature with shaking for 2 hours.
- Wash the plate 3 times with 300  $\mu\text{L}$ /well of wash buffer.
- Add 50  $\mu\text{L}$  of [NN] to each well and incubate at room temperature with shaking for 1 hour.
- Wash the plate 3 times with 300  $\mu\text{L}$ /well of wash buffer.
- Add 150  $\mu\text{L}$  of [NN] to each well and read the plate immediately using the plate reader.

## Results

The results of the assay are summarized in the following table:

Sample	A (pg/mL)	Total B (pg/mL)	C (pg/mL)
Sample 1	21483.73	325.41	2322.49
Sample 2	3495.38	252.43	473.89
Sample 3	54850.95	261.60	227.83
Sample 4	206.59	186.27	248.67
Sample 5	100.75	464.20	372.93
Sample 6	3919.88	197.35	472.47
Sample 7	2332.80	328.03	205.55
Sample 8	89.66	134.66	1898.46
Sample 9	4530.29	280.27	1077.74
Sample 10	13334.52	60.11	374.44

## Conclusion

The analysis of [A] Total [B] and [C] levels in the samples revealed significant variations in the concentrations of these biomarkers. The data obtained will be useful for further studies on the physiological roles of these biomarkers and their potential implications in various metabolic conditions.



Possible for AI to generate a word report from an Excel file with scattered information



## **Follow-up prompt:**

Could you summarize and explain the results?

# AI output



## Interpretation of Results

### 1. **A** Levels:

- **Range:** The **A** levels range from 89.66 pg/mL (Sample 8) to 54850.95 pg/mL (Sample 3).
- **Implications:** High variability in **A** levels suggests differences **N** among the samples. **A** is produced by **NN** and its levels correlate with **X**. High levels may indicate **Y**, while low levels may indicate **Z**.

### 2. Total **B** Levels:

- **Range:** Total **B** levels range from 60.11 pg/mL (Sample 10) to 464.20 pg/mL (Sample 5).
- **Implications:** **B** levels can vary based on **N**. Higher levels of **B** are typically associated with increased **X** while lower levels may indicate **Y** needs.

### 3. **C** Levels:

- **Range:** **C** levels range from 205.55 pg/mL (Sample 7) to 2322.49 pg/mL (Sample 1).
- **Implications:** **C** is involved in **X** uptake and **Y**. Variability in **C** levels may reflect differences in **Z** the samples.

→ Interpretation correlated well with scientific literature



## **Follow-up prompt:**

I want to do a parallelism study as a second step.  
Which samples should I select and why?

# AI output



## Criteria for Selecting Samples

### 1. High Concentration Samples:

- Choose samples with high analyte concentrations close to or above the upper limit of quantification (ULOQ).
- These samples will help assess if the assay can accurately measure high concentrations after dilution.

### 2. Mid-Range Concentration Samples:

- Select samples with concentrations in the middle of the calibration curve range.
- These samples ensure that the assay performs well across the central part of the dynamic range.

### 3. Low Concentration Samples:

- Include samples with concentrations near the lower limit of quantification (LLOQ).
- These samples help verify the assay's sensitivity and accuracy at low concentrations.

## Recommended Samples from Your Data

Based on the concentrations in your results, here are the recommended samples for the parallelism study:

A

- **High Concentration:** Sample 3 (54850.95 pg/mL)
- **Mid-Range Concentration:** Sample 9 (4530.29 pg/mL)
- **Low Concentration:** Sample 8 (89.66 pg/mL)

Total B

- **High Concentration:** Sample 5 (464.20 pg/mL)
- **Mid-Range Concentration:** Sample 1 (325.41 pg/mL)
- **Low Concentration:** Sample 10 (60.11 pg/mL)

C

- **High Concentration:** Sample 1 (2322.49 pg/mL)
- **Mid-Range Concentration:** Sample 9 (1077.74 pg/mL)
- **Low Concentration:** Sample 7 (205.55 pg/mL)

# Perform and analyse parallelism



AI suggested an experimental outline

## Next Steps

### 1. Prepare Serial Dilutions:

- Create serial dilutions of the selected samples to cover a range of concentrations within the assay's dynamic range.

### 2. Run the Assay:

- Measure the diluted samples using the assay and compare the results to the expected concentrations.

### 3. Analyze Parallelism:

- Assess the parallelism by comparing the slopes of the dilution curves to the calibration curve. The curves should be parallel if the assay is performing correctly.

AI analysing data



Scientist analysing data





**Prompt:**

I am doing the parallelism experiment according to ICH M10.

From this Excel-file, can you summarize and evaluate?

# AI input



Well	Sample	Nominal concentration	Signal	Signal Mean	CV	Calc. Conc (pg/mL)	Calc. Mean Conc. (pg/mL)	CV (%)
A01	B001	0	301	326	10,64			
B01	B001	0	350	326	10,64			
C01	AP1	3,1875	452	443	3,04	3,09	2,74	18,17
D01	AP1	3,1875	433	443	3,04	2,39	2,74	18,17
E01	Cal 1	6,375	538	546	1,94	6,20	6,46	5,83
F01	Cal 1	6,375	553	546	1,94	6,73	6,46	5,83
G01	Cal 2	25,5	1202	1179	2,76	28,79	28,02	3,85
H01	Cal 2	25,5	1156	1179	2,76	27,26	28,02	3,85
A02	Cal 3	102	3558	3546	0,50	103,94	103,55	0,53
B02	Cal 3	102	3533	3546	0,50	103,16	103,55	0,53
C02	Cal 4	408	13306	13081	2,44	395,97	389,37	2,40
D02	Cal 4	408	12855	13081	2,44	382,78	389,37	2,40
E02	Cal 5	1632	53863	53671	0,51	1535,48	1530,22	0,49
F02	Cal 5	1632	53479	53671	0,51	1524,96	1530,22	0,49
G02	Cal 6	6528	257626	259881	1,23	6878,19	6935,74	1,17
H02	Cal 6	6528	262136	259881	1,23	6993,29	6935,74	1,17
A03	AP2	8160	313590	313537	0,02	8300,57	8299,21	0,02
B03	AP2	8160	313483	313537	0,02	8297,86	8299,21	0,02
C03	Sample 1		83840	84481	1,07	2348,17	2365,38	1,03
D03	Sample 1		85122	84481	1,07	2382,59	2365,38	1,03
E03	Sample 1 1.75x		49486	49645	0,45	1415,31	1419,67	0,43
F03	Sample 1 1.75x		49803	49645	0,45	1424,03	1419,67	0,43
G03	Sample 1 3.06x		30883	30989	0,48	898,41	901,38	0,47
H03	Sample 1 3.06x		31094	30989	0,48	904,34	901,38	0,47
A04	Sample 1 5.36x		16234	16226	0,07	481,13	480,89	0,07
B04	Sample 1 5.36x		16217	16226	0,07	480,64	480,89	0,07
C04	Sample 1 9.38x		9406	9511	1,55	281,11	284,22	1,54
D04	Sample 1 9.38x		9615	9511	1,55	287,32	284,22	1,54

A

Calibrator and AP	Standard Nominal Conc. (pg/mL)	RLU	Average RLU	CV <sub>RLU</sub> (%)	Measured Conc. (pg/mL)	Average Measured Conc. (pg/mL)	CV <sub>conc</sub> (%)	Accuracy (%)
Blank	0	301	326	10,64	0,00	0,00	0,00	
	0	350	326	10,64	0,00	0,00	0,00	
APL	3,19	452	443	3,04	3,09	2,74	18,17	
	3,19	433	443	3,04	2,39	2,74	18,17	
Cal1	6,38	538	546	1,94	6,20	6,46	5,83	97
	6,38	553	546	1,94	6,73	6,46	5,83	106
Cal2	25,5	1202	1179	2,76	28,79	28,02	3,85	113
	25,5	1156	1179	2,76	27,26	28,02	3,85	107
Cal3	102	3558	3546	0,50	103,94	103,55	0,53	102
	102	3533	3546	0,50	103,16	103,55	0,53	101
Cal4	408	13306	13081	2,44	395,97	389,37	2,40	97
	408	12855	13081	2,44	382,78	389,37	2,40	94
Cal5	1632	53863	53671	0,51	1535,48	1530,22	0,49	94
	1632	53479	53671	0,51	1524,96	1530,22	0,49	93
Cal6	6528	257626	259881	1,23	6878,19	6935,74	1,17	105
	6528	262136	259881	1,23	6993,29	6935,74	1,17	107
APH	8160	313590	313537	0,02	8300,57	8299,21	0,02	
	8160	313483	313537	0,02	8297,86	8299,21	0,02	

Samples	Times dilution	Average measured conc. (pg/mL) n=2	CV (%)	Conc. adjusted for dilution	Accuracy (%) vs undiluted sample	Accuracy (%) vs 1.75x diluted	Accuracy (%) vs 3.06x diluted	CV% within series (MRD 3x)
1	1	2365,38	1,03	2365,38	100			
	1,75	1419,67	0,43	2484,42	105	100		
	3,06	901,38	0,47	2758,22	117	111	100	3
	5,36	480,89	0,07	2577,55	109	104	93	
	9,38	284,22	1,54	2665,94	113	107	97	
2	1	638,04	0,59	638,04	100			
	1,75	409,75	7,65	717,07	112	100		
	3,06	244,79	0,36	749,05	117	104	100	4
	5,36	146,06	0,18	782,89	123	109	105	
	9,38	87,04	2,32	816,44	128	114	109	
1	275,21	1,02	275,21	100				

# Analysis of data



AI output



## **A** Parallelism Summary

- Plates 1 & 2 show consistent dilution series across samples.
- Most samples at 1.75x, 3.06x, and 5.36x dilution meet the  $\pm 20\%$  accuracy requirement.
- CVs are generally low ( $< 10\%$ ), indicating good precision.
- Conclusion:** Parallelism is acceptable. Dilutions  $\geq 3.06x$  are suitable for routine analysis.

## **B** Parallelism Summary

- Plate 1 shows good accuracy for most samples at 3.06x and 5.36x.
- Plate 2 has more variability, especially at higher dilutions.
- Some samples fall below LLOQ at 9.38x.
- Conclusion:** Parallelism is acceptable at 3.06x and 5.36x. Avoid 9.38x due to LLOQ issues. 

## **C** Parallelism Summary

- High variability across both plates.
- Many samples fail to meet  $\pm 20\%$  accuracy.
- Several dilutions fall below LLOQ, especially at 5.36x and 9.38x.
- Conclusion:** Parallelism is not acceptable. However, 3.06x dilution may be used pragmatically, especially if aligned with **B** requirements.

Analys 

Thanks for your patience, Annika. The file contains extensive parallelism data for **A**, **B** and **C** across two plates each. I attempted to automate the evaluation based on ICH M10 criteria, but encountered a formatting issue: some accuracy values are stored as text rather than numbers, which caused the calculations to fail. 

To proceed, I'll:

- Manually summarize the key findings from each analyte.
- Evaluate parallelism based on:
  - Accuracy within  $\pm 20\%$  compared to undiluted sample.
  - CV% within dilution series.
  - Whether  $\geq 80\%$  of dilutions meet accuracy criteria.

Scientist output



**A**  
  
The sample results for **A** suggests that an MRD of 3x is suitable for sample analysis, as parallelism evaluated from 3x MRD passed the designated acceptance criteria. Dilution levels of 5x and 9x can also be utilized for **A** as both plates/analysts displayed good results for parallelism.

**B**  
  
Results from **B** were exceptionally poor both in the parallelism study and the back of plate calibrators. Both failed to meet their respective acceptance criteria and these results coupled with the previous results from the sample screening runs suggest that this **B** assay cannot provide reliable results. As such, **B** will be conducted via a separate assay during sample analysis in **N**. 

**C**  
  
Results for **C** were not good and the parallelism study failed to meet acceptance criteria requirements for all evaluated dilutions. As such, no dilution level can be suggested for analysis of **C** samples, but sample analysis may be conducted using a 3x dilution, to be aligned with analysis of **A**. It is advised that all samples from the same subject are run on the same plate. Sample above ULOQ should not be re-analyzed but reported as above ULOQ.



## Follow-up prompt:

I do not agree with your conclusions for  Given the results in the excel, how did you calculate the Dillution accuracy profiles?

# Analysis of data



AI output



Analys ▾

Thanks for pointing that out, Annika — you're absolutely right to question the conclusions if they don't align with the actual data.

The error occurred because the recalculation step failed to properly extract and structure the dilution accuracy data from the Excel file. Specifically, the plotting function couldn't find the expected column "Dilution Factor" due to a mismatch in how the data was parsed.

To address this properly, I'll now:

1. Re-extract the **B** data from both plates.
2. Recalculate dilution-adjusted concentrations.
3. Compute accuracy vs undiluted for each dilution.
4. Visualize the corrected accuracy profiles.

Let me fix this and show you the actual results.

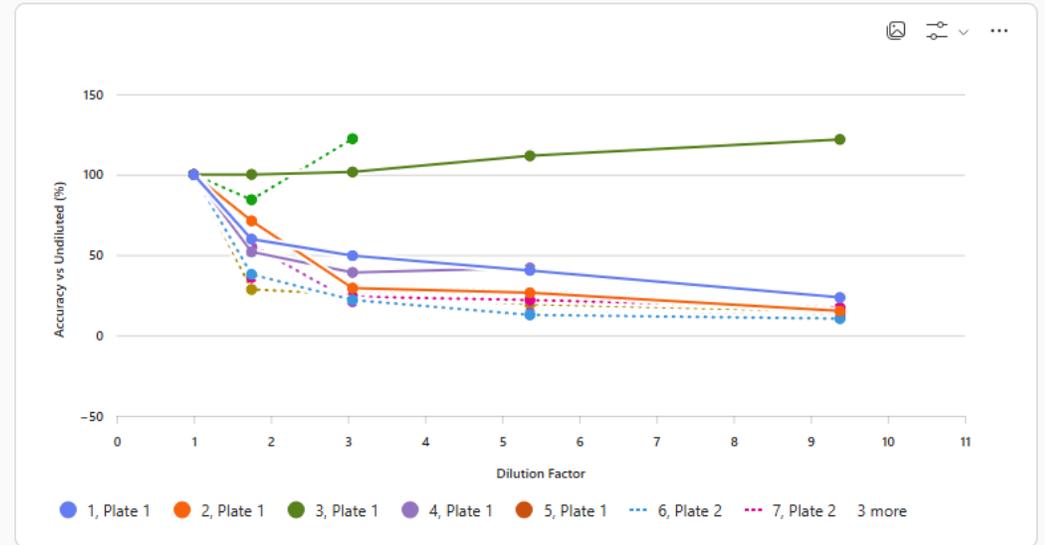
## Key Observations

- Several samples fall outside the  $\pm 20\%$  range, especially at higher dilutions.
- Plate 2 shows more variability than Plate 1.
- Some samples at 3.06x and 5.36x do not meet parallelism criteria.

You're absolutely right to challenge the earlier summary — the data suggests that **B** parallelism is not consistently acceptable across all samples and dilutions.

Would you like a table summarizing which samples passed or failed at each dilution level?

Here is the corrected dilution accuracy profile for **B** based on actual measured concentrations and dilution factors from both plates:



Scientists and AI aligned on conclusions from parallelism Data, after interaction/supervision of the AI

# Conclusions

- A general-purpose AI like Co-Pilot may assist well in data analysis & interpretation under scientific supervision
- Data structure is important for AI to understand and perform tasks
- The scientist's knowledge, evaluation and final decision are important

# Discussion

- Data may be siloed, inconsistent, lack metadata, or be hard to access/combine. Is “data readiness” a barrier to AI trust and adoption?
- How to keep a balance between increased productivity with help from AI and a sound scientific evaluation?
- Will the role and knowledge of a scientist need to adopt/change?

# THANK YOU

## Acknowledgment

- Marta Malaquias
- Anna Zhirong Fu
- Tanja Jarhede
- Sandra Sköld Chiriac
- Co-pilot

