

# A Comparative Analysis of LLMs' Responses for governance dimension

## Quantitative Indicators :

**Indicator : G1.1.** Pourcentage de femmes dans la gouvernance.

**Question : G1.1** – Quel est le pourcentage de femmes au sein des instances de gouvernance et de direction de l'organisation ?

Poulina : All the LLMs succeeded in giving the right pourcentage . DeepSeek and ChatGPT gave more details , ChatGPT added the percentage of men , and DeepSeek gave the percentage of women and the total number in the governance and management bodies of the organization..

SFBT :Only DeepSeek and Copilot succeed in extracting information about the number of women in the board of directors .

==> It is important to clarify more this indicator to the LLM( The percentage of women in the board of directors should be included in this indicator , and explain the difference between G1.1 and G1.2. )

Délice :All the LLMs indicated that the answer to this question was not present in the ESG report ,except ,Copilot , gave a wrong answer .

UBCI : Only Copilot provided a detailed answer , and Gemini failed to answer this question .

Assurance Maghreb \_vie , Assurance Magherbia : All the LLMs extracted the percentage , except Gemini .

Assurance Star : In the ESG report , the number of women and men was mentioned , all the LLMs succeeded in extracting this information and calculating the rate of women in the governance and management bodies of the organization by themselves .

TLF : All the LLMs indicated the absence of this rate .

**Indicator : G1.2.** Pourcentage de femmes dans les comités du CA

**Question : G1.2.** Quel est le pourcentage de femmes siégeant dans les comités du Conseil d'Administration ?

Poulina : All the LLMs provided the right percentage .

SFBT: Copilot fails to distinguish between the board of committees and the board of directors ( it exhibited confusion between G1.1 and G1.2.)

Délice : All the LLMs mentioned the absence of this percentage , except Copilot provided a wrong percentage .

UBCI : Only ChatGPT was able to distinguish between the board committees and the board of directors.

Assurance Maghreb vie , Assurance Maghreb : all the LLMs provided the true percentage except Gemini .

Assurance Star : In Assurance Star 's report , the percentage of women on board committees was mentioned implicitly ( the number of women and men on board committees existed but not the percentage ). Copilot measured the percentage by itself , Gemini provided only the numbers , DeepSeek delivered a wrong percentage and ChatGPT mentioned the absence of this information .

⇒ We can add in the prompt how to measure the rate if only the number of men and women is mentioned .

TLF : Délice : All the LLMs mentioned the absence of this percentage

**Indicator : G2.2.** Pourcentage d'administrateurs indépendants :

**Question : G2.2 .** Quel est le pourcentage d'administrateurs indépendants au sein du Conseil d'Administration ?

Poulina , Délice ,TLF : All the LLMs indicated that the answer to this question was not present in the ESG report .

SFBT : Copilot and Gemini gave the number of independent directors on the board of directors when they didn't find the percentage .While ChatGPT gave the answer "Non chiffré " and DeepSeek gave the number that existed in the report and calculated the percentage by itself.

UBCI , Assurance Maghreb vie ,Assurance Maghreb : : All LLMs succeeded in extracting the percentage except Gemini

Assurance Maghreb : Copilot and DeepSeek succeeded in extracting the percentage ,

Assurance Star : All the LLMs gave the true percentage .

**Indicator : G4.** Pourcentage d'employés couverts par des accords collectifs.

**Question : G4 –** Quel est le pourcentage total d'employés couverts par des accords de négociation collective ?

Poulina : All the LLMs succeeded in giving the right percentage .

SFBT ,Délice,TLF : All the LLMs indicated the absence of this percentage in the ESG report .

UBCI , Assurance\_Maghreb\_vie , Assurance Maghreb,Assurance Star : All the LLMs succeeded in giving the right percentage,except Gemini .

### **G5.3. Pourcentage de fournisseurs analysés selon critères sociaux/environnementaux**

Question : G5.3 – Quel pourcentage de fournisseurs et de partenaires de la chaîne d’approvisionnement ont été analysés selon des critères sociaux et environnementaux ? Quels sont ces critères ?

Poulina : All the LLMs when they didn’t find the percentage of suppliers , they extract their numbers , except Copilot.

**⇒ It is crucial to enhance the difference between the number and the percentage in this situation .**

SFBT : All the LLMs mentioned the absence of this information ,except Gemini that gave wrong answer.

### **Indicator : G6.2. Pourcentage de personnes formées à l’éthique/anticorruption.**

**Question : G6.2** – Quel pourcentage des membres de la gouvernance, des employés et des partenaires ont été formés aux questions éthiques et aux politiques/procédures anti-corruption ?

Poulina : In this report all the members of the governance and the employees had been trained in ethical issues , but an exact percentage wasn’t mentioned , all the LLMs succeeded in extracting this information except Gemini.

SFBT : In this report all the members of the governance and the employees had been trained in ethical issues , but an exact percentage wasn’t mentioned , all the LLMs succeeded in extracting this information except ChatGPT.

**==> we can add in the prompt if all the employees had been trained , the LLM should extract this information and write 100% .**

Délice , Assurance Maghreb \_vie , TLF : All the LLMs succeeded in highlighting the absence of any percentage .

UBCI : All the LLMs extracted the percentage from the report except Gemini .

Assurance Maghreb \_vie : All the LLMs succeeded in highlighting the absence of any percentage , except ChatGPT that provided a wrong answer .

Assurance Star : Each LLM gave a different answer , in this report a number of employees was mentioned and not a specific percentage : Copilot gave a wrong percentage , Gemini and ChatGPT mentioned the absence of this percentage and DeepSeek provided the percentage of the participants that have successfully completed the training .

### **Indicator : G8. Taux de départements évalués pour conformité réglementaire:**

Question : G8 – Quel est le taux des départements internes ayant été soumis à une évaluation de conformité réglementaire ?

Poulina : In the report all the departments had been affected by a regulatory compliance assessment. But no direct percentage was mentioned , all the LLMs succeeded in extracting this information except Copilot .  
==>we can add in the prompt if all the departments had been affected , the LLM should extract this information and write 100%

SFBT, Délice, Assurance Maghreb vie , TLF : All the LLMs succeeded in highlighting the absence of any rate.

UBCI : All the LLMs succeeded in highlighting the absence of any percentage , except Gemini.

Assurance Maghreb : All the LLMs succeeded in highlighting the absence of any percentage, except ChatGPT , that mentioned a percentage of 100%.

Assurance Star : Gemini and ChatGPT failed to extract this rate .

## Qualitative Indicators :

### **Indicator G2.1.** Séparation des fonctions DG / Président (Oui/Non)

Question : G2.1 . Y a-t-il une séparation entre les fonctions de Directeur Général et celles de Président du Conseil d'Administration ?

Poulina ,SFBT, UBCI and Star insurance : All the LLMs succeeded at answering this question .

Délice : All the LLMs indicated that the answer to this question was not present in the ESG report ,except ,Copilot , gave a wrong answer .

Assurance Maghreb vie and Assurance Maghreb : All the LLMs succeed in extracting information about the separation between the functions of the chief executive officer and that of the chairman of the board of directors ,except ,Gemini .

TLF , All the LLMs indicated that the answer to this question was not present in the ESG report

### **Indicator : G3.1.** Publication de la politique de rémunération (Oui/Non)

Question : G3.1. L'entreprise publie-t-elle des informations sur les politiques de rémunération en vigueur au sein de l'organe de direction ?

Poulina , Assurance\_Maghreb\_Vie , Assurance Maghreb : All the LLMs succeeded in answering this question.

SFBT : All the LLMs indicated that the answer to this question was not present in the ESG report ,except ChatGPT , that gave a wrong answer.

Délice : All the LLMs indicated that the answer to this question was not present in the ESG report ,except Copilot , that gave a wrong answer.

UBCI : All the LLMs indicated that the answer to this question was not present in the ESG report ,except Gemini , that gave a wrong answer.

Assurance Star : Only Gemini and ChatGPT succeed to extract this information .

TLF : All the LLMs indicated that the answer to this question was not present in the ESG report , but the HR policy was mentioned without details , except Gemini.

**Indicator : G3.2.** Critères RSE dans la rémunération (Oui/Non)

**Question :** G3.2 . Des critères de performance RSE sont-ils pris en compte dans la rémunération des cadres dirigeants ?

Poulina : All the LLMs effectively retrieved pertinent answers .

SFBT,Assurance\_Maghrebia\_Vie,TLF: All the LLMs mentioned that this report does not include any information related to this indicator.

Délice ,Assurance Star : All the LLMs mentioned the absence of any information related to this indicator in this report . , except Copilot.

UBCI : All the LLMs mentioned the absence of any information related to this indicator in this report . , except Gemini.

Assurance Mghrebia : All the LLMs mentioned the absence of any information related to this indicator in this report . , except Copilot.

**Indicator : G5.1.** Description de la chaîne d’approvisionnement

**Question :** G5.1.Comment la chaîne d’approvisionnement de l’organisation est-elle décrite ?

Poulina : All the LLMs gave answers to this question .We find that DeepSeek provides the most detailed answer and cites the page number from the report .

SFBT : All the LLMs answer this answer correctly in different ways , except ChatGPT that gave a binary answer.

Délice : All the LLMs gave answers to this question .We find that ChatGPT provides the most detailed answer .

UBCI : All the LLMs provide answers to this question .We find that Copilot gave the best answer .

Assurance Maghreb vie : All the LLMs gave answers to this question ,except DeepSeek.

Assurance Maghreb : Only DeepSeek mentioned that there is no detailed information about this question , the other LLMs provide wrong answers.

Assurance star : All the LLMs provided answers to this question except Gemini .

TLF : Each LLM provide different answer

.==> **We have to detail the meaning of the supply chain for the financial sector in the prompt .**

#### **Indicator : G5.2.** Systèmes et critères d'analyse des fournisseurs

Question :G5.2 – Quels systèmes sont utilisés pour analyser les fournisseurs selon des critères sociaux et environnementaux ? Quelle est la liste de ces critères ?

Poulina , Délice ,UBCI , Assurance Maghreb vie : All the LLMs provide answers to this indicator .

SFBT : All the LLMs mentioned that this report does not include any information related to this indicator, except Copilot.

Assurance Maghreb : All the LLMs provided answers to this indicator that needed to be more developed .

Assurance Star : All the LLMs provide the correct answer except Gemini.

TLF : We got different answers.

#### **Indicator :G6.1.** Existence d'un code de conduite/éthique (Oui/Non)

Question : G6.1 – L'entreprise dispose-t-elle d'un code de bonne conduite, de déontologie et/ou d'éthique ?

Poulina : All the LLMs succeeded in extracting information about this indicator , but ChatGPT gave the most detailed answer .

SFBT ,UBCI ,Assurance Maghreb-vie ,Assurance Maghreb ,Assurance Star ,TLF: All the LLMs mentioned the existence of this code .

Délice : All the model successfully extracted indirect information , except Gemini .

**Indicator : G6.3.** Existence d'une politique anti-corruption (Oui/Non)

**Question : G6.3** – L'entreprise dispose-t-elle d'une politique de lutte contre la corruption ?

Poulina , Assurance Maghreb vie : All the LLMs succeed in citing the presence of an anti-corruption policy , except Gemini .

SFBT, UBCI, Assurance Maghreb, Assurance Star , TLF: All the LLMs succeed in citing the presence of an anti-corruption policy .

Délice : Only ChatGPT and Copilot succeeded in extracting implicitly expressed information.

**Indicator : G7.** Existence d'une politique de confidentialité des données (Oui/Non).

**Question : G7** – L'entreprise dispose-t-elle d'une politique de confidentialité des données ?

Poulina : All the LLMs succeed in citing the presence of a data privacy policy, except Gemini .  
SFBT, UBCI , Assurance Maghreb vie, Assurance Maghreb , Assurance Star, TLF: All the LLMs succeed in citing the presence of a data privacy policy.

Délice : All the LLMs succeed in citing the absence of a data privacy policy , except Copilot

**Indicator : G9.1.** Liste des parties prenantes engagées.

**Question : G9.1** – Quels sont les groupes de parties prenantes avec lesquels l'entreprise a engagé un dialogue ?

Poulina : The list of stakeholder groups was mentioned implicitly in the ESG report , all the LLMs succeeded in extracting it , except Gemini.

SFBT: The list of stakeholder groups is presented in the ESG report in the form of an image , all the LLMs succeeded in extracting it , except Gemini.

Délice , UBCI , TLF: All the LLMs succeeded in extracting the stakeholder's list .

Assurance Maghreb vie , Assurance Maghreb , assurance Star : All the LLMs succeeded in extracting the stakeholder's list except Gemini .

**Indicator : G9.2.** Modalités et fréquence du dialogue :

**Question : G9.2** – Quelles sont les modalités et la fréquence du dialogue selon le type et le groupe de parties prenantes ?

Poulina : Gemini and ChatGPT failed in extracting information related to this indicator .

SFBT, Délice : All the LLMs succeeded in extracting information related to this indicator .

UBCI : All the LLMs indicate that the information related to this indicator is not detailed in the report .

Assurance Maghreb vie , assurance Maghreb , Assurance Star: All the LLMs succeeded in extracting information related to this indicator , except Gemini .

TLF : Gemini and DeepSeek failed in extracting information related to this indicator .

**Indicator : G9.3.** Questions soulevées et réponses de l'organisation.

**Question : G9.3** – Quelles ont été les questions ou préoccupations clés soulevées lors de ces dialogues, et comment l'organisation y a-t-elle répondu ?

Poulina : Only Copilot succeeded in extracting information related to this indicator .

SFBT : ChatGPT and Copilot gave wrong answers to this question.

Délice ,UBCI : All the LLMs succeeded in extracting information about this indicator with different answers .

Assurance Maghreb vie , Assurance Maghreb , Assurance Star: All the LLMs succeeded in extracting information about this indicator with different answers , except Gemini .

TLF : All the LLMs mentioned the absence of any information related to this indicator in the report .

**Indicator : G10.1.** Part des achats auprès des startups ou initiatives en faveur des startups.

**Questions : G10.1** – Quelle part des achats de l'entreprise est réalisée auprès de startups ou dans le cadre d'initiatives en leur faveur ?

Poulina : All the LLMs mentioned the presence of some information related to this indicator without rate.

SFBT , Assurance Star, TLF: All the LLMs mentioned the absence of any information related to this indicator in the report .

Délice : all the LLMs succeeded in extracting the percentage related to this indicator .

UBCI : All the LLMs mentioned the absence of any information related to this indicator in the report , except ChatGPT mentioned the existence of an initiative.

Assurance Maghreb vie : Copilot and Gemini fail to extract the rate of purchases from startups from the Assurance Maghreb vie report .



Assurance Maghreb : Gemini fails to extract the rate of purchases from startups from the Assurance Maghreb report .

**Indicator : G10.2.** Part des achats auprès de TPE, avec indication du pourcentage de coopérative et de toute entreprise de l'ESS (Economie Sociale et Solidaire).

**Question : G10.2** – Quelle part des achats est réalisée auprès de TPE ? Quel est le pourcent

Poulina : All the LLMs succeeded to extract information related to this indicator from the ESG report 0

SFBT , Délice ,TLF : All the LLMs mentioned the absence of any information related to this indicator in the report.

UBCI : Only ChatGPT provided a good answer.

Assurance Maghreb\_Vie : Copilot gave a wrong answer about this indicator .

Assurance Maghreb : Each LLM gave a different answer .

Assurance Star : All the LLMs mentioned the absence of any information related to this indicator in the report , except DeepSeek.

**Indicator : G11.** Publication d'un rapport RSE-DD (Oui/Non).

**Question : G11** – L'entreprise publie-t-elle un rapport RSE ou de développement durable ?.

All the LLMs succeeded in mentioning the presence of the ESG report for all the companies .

## Important Findings from the Analysis

The analyses of this indicators showed that , Gemini presents the worst LLM among others in extracting informations related to social indicators .DeepSeek , Copilot and ChatGPT demonstrated better performance , but they failed sometimes in extracting the right information , give wrong answers or make confusion between two indicators especially ChatGPT .Demonstrating that quantitative indicators need more detailed prompts.

