A Comparative Analysis
of LLMs' responses for Social dimension

Quantitative Indicators:

Indicator : S1. Le ratio entre la rémunération moyenne des hommes et la rémunération moyenne des femmes et ce par catégorie professionnelle.
Question : S1. Quel est le ratio entre la rémunération moyenne des hommes et celle des femmes, par catégorie professionnelle ?

Poulina: Gemini fails to extract the information from the ESG report. DeepSeek and ChatGPT gave more detailed information thanCopilot (Even DeepSeek mentioned the page number from which it extracted the data).

UBCI: When DeepSeek , Gemini and Chatgpt didn't find the answer of the question in the ESG reports, they mentioned that the information did not exist, whereas Copilot skips the question and proceeds to the next one.

Assurance Maghrebia_vie : Even though the information exists and is detailed , Gemini fails to extract it .
Assurance Maghrebia : Even though the information exists and detailed , Gemini and ChatGPT fail to extract it.

⇒**For the indicator S1, Gemini fails to extract information from the 8 companies' reports.**

Indicator : S2.1. Répartition des effectifs par type de contrat (CDI, CDD, à plein temps, en temps partiel, en télétravail).
Question:S2.1.Quelle est la répartition des effectifs selon le type de contrat (CDI, CDD, plein temps, temps partiel, télétravail) ?

Poulina: ChatGpt and DeepSeek gave the best results , in contrast to Gemini that fails to extract the information .

Assurance_Maghrebia vie : Even though the information exists and is detailed , Gemini fails to extract it.

Assurance Maghrebia : Even though the information exists and is detailed , Gemini fails to extract it , and ChatGPT provides wrong answers.

Indicator : S2.2. Nombre de contrats d'insertion (apprentissage, par alternance, karama, civp,. . .)conclus dans l'année et pourcentage des contrats convertis en contrats CDI.
Question: S2.2. Combien de contrats d'insertion (apprentissage, alternance, Karama, CIVP, etc.) ont été conclus durant l'année, et quel est le pourcentage de ces contrats convertis en CDI ?

Poulina and Délice : Gemini fails to extract the information, although it is very detailed.

UBCI : Gemini and DeepSeek mentioned the number of interns as an integration contract.
⇒ **It is important to add in the prompt that the interns ' number is not included.**

Assurance Maghrebia : Even though the information exists and is detailed , Gemini fails to extract it , and ChatGPT gave wrong answers.

Assurance Star : All the LLMs fail to extract the information because , it is presented in the form of images.

Indicator :S2.3 La rotation des effectifs par type de contrat (CDI, CDD, à plein temps, en temps partiel) d'une année à l'autre.
Question :S2.3 Quelle est la rotation des effectifs par type de contrat (CDI, CDD, plein temps, temps partiel) d'une année à l'autre ?

Poulina Group Holding : the stuff turnover by type of contact presented in the ESG report in an histogram , Only Copilot succeed in extracting it .

Assurance Maghrebia : Even though , the information exists and detailed , Gemini fails to extract it , and ChatGPT gave wrong answers.

Assurance Star : the stuff turnover by type of contact presented in the ESG report in an image , Only Copilot succeed in extracting it, and DeepSeek gave a wrong answer .

Indicator :S3. Répartition hommes/femmes par catégorie professionnelle.
Question : S3. Quelle est la rotation des effectifs par type de contrat (CDI, CDD, plein temps, temps partiel) d'une année à l'autre ?

Poulina Group Holding : Gemini couldn't extract the information and DeepSeek failed to provide complete information .

Assurance Maghrebia_vie : Gemini couldn't extract the information and Copilot failed to provide complete information

Assurance Star : This information related to this indicator is presented in an image , which make it difficult for some LLMs to extract it . Only Copilot and ChatGPT succeed in extracting it .

Indicator : S5.2 Le taux d'accidents de travail (TAT), le taux de maladies profession nelles (TMP), le taux de journées de travail perdues (TJP), le taux d'absentéisme (TA) et les décès liés au travail pour tous les employés, avec une répartition par genre.
Question : S5.2 Quels sont les taux suivants pour l'ensemble des employés (avec une répartition par genre) : Taux d'accidents de travail (TAT), Taux de maladies professionnelles (TMP), Taux de journées de travail perdues (TJP), Taux d'absentéisme (TA),Nombre de décès liés au travail ?

Poulina Group Holding : Gemini was unable to extract the information.

SFBT: Gemini was unable to extract the information, ChatGPT gave the information of 2022, Copilot gave the best answer.

==> It is essential to mention the year with the indicator , because some companies show evolutions of their activities across time.

Délice: Gemini was unable to extract the information , ChatGPT gave the best answer .

Assurance _Maghrebia : Even though , the information exists and detailed , Gemini fails to extract it , and ChatGPT gave wrong answers.

Assurance Star : Gemini was unable to extract the information .

Indicator :S9.1. Nombre moyen d'heures de formation par an, par salarié et par catégorie professionnelle .
Question :S9.1. Quel est le nombre moyen d'heures de formation par an, par salarié et par catégorie professionnelle ?

Poulina , SFBT, Délice : Gemini was unable to extract the information .

Assurance Maghrebia : Only Copilot was able to give the correct answer .

Indicator : S9.2. Nombre moyen d'heures de formation dédiées aux thèmes environnementaux et sociétaux.
Question:S9.2. Quel est le nombre moyen d'heures de formation consacrées aux thématiques environnementales et sociétales ?

Assurance Maghrebia : DeepSeek could not extract data , and ChatGPT gave wrong information .

Indicator :S10.2. Le pourcentage du chiffre d'affaires de l'entreprise investit au niveau de la communauté locale.
Question: S10.2. Quel pourcentage du chiffre d'affaires de l'entreprise est investi dans la communauté locale ?

UBCI: Copilot could not extract partial information about the total capital invested.

Assurance Maghrebia:ChatGPT gave wrong answers.

Qualitative Indicators:

Indicator : S4 . Existence d'une charte ou d'une politique de la diversité et de non-discri mination (Oui/Non)
Question : S4 . L'entreprise dispose-t-elle d'une charte ou d'une politique de diversité et de non-discrimination ? (Oui/Non).

Poulina Group Holding : Gemini could not extract the data .

Délice : Copilot could not extract the data .

Indicator : S5.1 Liste des types d' accidents de travail et de maladies professionnels.
Question : S5.1 Quels sont les types d'accidents de travail et de maladies professionnelles recensés dans l'entreprise ?

Délice : The list of type of accidents and occupational diseases didn't exist in the ESG report , Copilot et DeepSeek indicate the the rate of work accidents in their answers , which are related to the next indicator(S5.2.)

⇒**We have to mention in the prompt the difference between the two indicators**

indicator :S6. Analyse des risques liés à la santé et à la sécurité au travail (SST) et mise en place d'un plan d'atténuation des risques SST y compris les risques psychosociaux (Oui/Non).
Question : S6. Une analyse des risques liés à la santé et sécurité au travail (SST), incluant les risques psychosociaux, a-t-elle été réalisée ? Un plan d'atténuation des risques a-t-il été mis en place ? (Oui/Non)

Poulina Group Holding : DeepSeek and ChatGPT made the best answers.They are more detailed than the response provided manually.

TLF :Only ChatGPT was able to mention the strategic initiative about the security at work.

indicator :S7.1. L'existence d'une politique destinée à l'élimination (abolition) de toute forme de travail forcé et/ou des enfants(Oui/Non).
Question : S7.1. L'entreprise dispose-t-elle d'une politique visant à éliminer toute forme de travail forcé ou d'enfants ? (Oui/Non).

Poulina : Gemini was unable to extract the information.

SFBT : Copilot was unable to detect the initiative of SFBT of all forms of forced children labor.

Délice : ChatGPT provided the information from human rights , which is related to the indicator (S8.1)

⇒ **We must mention in the prompt the difference of the two indicators.(s7.1 and s8.1).**

TLF : The response of DeepSeek and ChatGPT closely aligns with the manual response by mentioning the implicit commitment about the forced child labor. In contrast , Gemini mentions the absence of direct information. Gemini and copolite mentioned the
absence of this information.

Indicator :S7.2. Si oui, est ce que cette politique est communiquée aux fournisseurs et aux clients (Oui/Non).
Questions: S7.2. Si oui, cette politique est-elle communiquée aux fournisseurs et aux clients ? (Oui/Non)

Délice : We must mention in the prompt the difference of the two indicators.(s7.1 and s8.1), because ChatGPT used the information about human rights as a response of S7.1.

TLF : Only ChatGPT succeeded to answer this question.

Indicator: S8.1. L'existence d'une politique aux sein de l'entreprise relative au droits de l'Homme (Oui/ Non)
Questions : S8.1.L'entreprise a-t-elle une politique relative aux droits de l'Homme ? (Oui/Non).

Délice : ChatGPT excels in extracting explicit data, even surpassing all other models and the response provided manually.

Indicator : S.8.2 Si Oui, Est-ce que cette politique couvre les clients et les fournisseurs (Oui/ non).
Questions :S.2. Si Oui, Est-ce que cette politique couvre les clients et les fournisseurs (Oui/ non ?

Délice :Gemini et ChatGPT excel in extracting partial data, surpassing the other models and manually extracted information .

TLF : Only ChatGPT succeeded to answer this question.

Indicator ; S10.1 Liste des programmes de développement des communautés locales fondés sur leurs besoins.
Questions : S10.1.Quels sont les programmes de développement des communautés locales mis en place par l'entreprise, fondés sur leurs besoins ?

Poulina : ChatGPT asks for more details about this indicator to be able to answer the question .

Assurance Maghrebia :chatGPT gave wrong answers .

Important Findings from the Analysis

Gemini fails to extract data related to the social indicators from ESG reports in most cases.
And ChatGPT frequently gave wrong information, especially for Maghrebia Assurance and excels at extracting information better than other LLMs in other cases , it gives wrong answers for all the indicators.Moreover, all the LLMs commonly fail to extract data from assurance Star 's ESG report (The structure of the report is different from the other reports , the data and percentage are generally presented in the form of an image).
Besides , the LLM occasionally gave more detailed information than the manual responses .DeepSeek and Copilot provided better results for social dimensions than ChatGPT and Gemini , although they failed in certain cases.