

GIST Expert Annotation Guidelines

Guideline Structure

1. Introduction
2. Annotations
3. Recommendations
4. Examples for typical errors/difficulties

Introduction

Intro

- This slidedeck guides you through the manual expert annotation of **Scope 1/2/3 greenhouse gas emissions** in company reports.
- Please read through the **entire document** before starting the annotation.
- In case of unclarities please use the “Comment” columns in the annotation files

Intro

- You will enter your annotations in an Excel file
- Each Excel file corresponds to exactly one company report
- We do not prescribe in what particular order you fill in your Excel sheet for one report

Intro

- We retrieve the annotations from you (experts) in order to
 1. Get the correct, gold-standard annotations
 2. With that assess the annotation quality of LLMs and non-experts
 3. Train new models/create new pipelines for high-quality extraction
 4. Find out which annotation setup works better (Experimental Variation)

Annotations

Basic Annotation Setup

- During your annotations you will see what the non-expert humans annotated
- You will work in groups of 2
- Each report will be annotated twice, once per group
- The idea is that you meet with the corresponding other expert and discuss the report together to come up with the gold standard
- You will work with every other expert on ~20 reports each
- In case of disagreement between the two expert pairs it will come to a second iteration of discussion
- Out of 140 reports, roughly 60 require expert adjudication
(more details later in this slidedeck)

Reports selected for expert adjudication

So far all 140 reports were double-annotated by 2 non-experts. Many of them were empty, on some the non-experts agreed.

Reports were selected for expert annotation if:

1. The non-experts requested it
 - a. For a whole document
 - b. For one row (= one Scope-Year combination)
2. The non-experts disagreed in their annotation for
 - a. The emission value and/or
 - b. The emission unit

vitalenergyinc 2019 report.pdf

The rows of interest that we need you to annotate are highlighted with color!

[illegible]

vital energy
inc_2019_re
port.pdf

If expert was requested for the whole document, it is indicated at the top of the document. In that case, all rows with LLM-extracted values need to be annotated (yellow or not)

ID	Page used	LLM Year	LLM Score	(yellow or not)												Additional Comment (if helpful)	Unclear record (Yes/No)	
							(Yes/No)	necessary)	choose first)		(Yes/No)	necessary)	apply choose first)		(Yes/No)	necessary)		
12																		
13			2013	1			n/a				n/a				Yes			No
14			2014	1			n/a				n/a				Yes			No
15			2015	1			n/a				n/a				Yes			No
16			2016	1			n/a				n/a				Yes			No
17	325	15	2017	1		63.3	million metric tons				n/a				Yes			No
18	326	66	2017	1		63.3	million metric tons				n/a				Yes			No
19	327	15	2018	1		59.5	million metric tons				n/a				Yes			No
20	328	66	2018	1		59.5	million metric tons				n/a				Yes			No
21	329	15	2019	1		47	million metric tons				n/a				Yes			No
22	330	66	2019	1		47	million metric tons				n/a				Yes			No
23			2020	1							n/a				Yes			No
24			2021	1							n/a				Yes			No
25			2022	1							n/a				Yes			No

General Info

[illegible]

- **Focus on the highlighted rows**
 - If no row is highlighted, expert adjudication was requested for the entire document ⇒ Check the document-level comments
- **Focus on the non-gray columns**
 - Gray columns are values extracted by the LLM or annotated by lay annotators and must not be changed
 - Your cells have white background

Report-level information

1				
2	Report Name	blackberry ltd_2022_report.pdf		
3	Document comment Annotator 1			
4	Document comment Annotator 2			
5	Document-level expert adjudication requested by at least 1 annotator	No		
6	Expert ID 1			
7	Expert ID 2			
8	Expert document comment			
9	Annotation started			
10	Annotation ended			
11				

- **Report name:** the report's name (do not change)
- **Document Comment 1:** document-level comment of lay annotator 1
- **Document Comment 2:** document-level comment of lay annotator 2
- **Document-level expert adjudication requested by at least 1 lay annotator:** If this cell is red, a non-expert requested the expert to have a look at the whole document + report (do not change)
- **Expert ID1 and Expert ID2:** Please put in your IDs
- **Expert doc. Comment:** Open field for comment.
 - Please use this to explain document-level decisions/comments, document unclarities and/or open questions. Your input is very valuable
- **Annotation started:** Fill in timestamp (HH:MM) when you started the annotation
- **Annotation ended:** Fill in timestamp (HH:MM) when you ended the annotation
- All cells **highlighted in red** are mandatory ⇒ please check that you filled these cells

Starting the expert annotation

1. Open the Excel and the corresponding company report
2. Share your screen for your fellow expert
3. Check the time and fill Annotation started timestamp
4. Enter Expert ID 1 & 2
5. Check whether document-level expert adjudication requested by at least 1 annotator
 - a. if yes \Rightarrow check document comment Annotator 1 & 2, also annotate all rows with LLM-extracted values
 - b. if no \Rightarrow continue
6. Check out highlighted rows \Rightarrow Walk through column by column

When is a value correct?

To be a correct value of interest it needs to fulfill the following criteria:

1. It covers the emissions of the **whole company** (in accounting language “whole company” is sometimes called “consolidated”)
 - a. not e.g., just supply chain or just facilities in Bangladesh
2. The emissions are reported according to the operational boundaries of the **Scopes** (or other valid categorizations like direct / indirect)
 - a. no “custom” boundaries like supply chain emissions or total emissions, net emissions etc. are introduced
3. The company reports **absolute GHG** (mostly CO2 equivalent, sometimes only CO2) emissions
 - a. not e.g., SO2 emissions or CO2 emissions per passenger

Every single value that does not meet all of these criteria is always to be annotated as wrong! Note that empty cells can be wrong when there is a matching year-scope value in the report that has not been extracted.

Every wrong value needs to be reasoned why wrong (→ next slide)

Reasoning

The categorical reasoning variables can take the following values:

Value Reasoning	U
(select if necessary, if multiple apply choose first)	cc (Y
<div>▼</div>	
0. Missed out on correct value (false NA)	n/
1. WV: Irrelevant, not absolute GHG emissions	n/
2. WV: Extracted value is not related to the whole company	n/
3. WV: Extracted value is related to different scope	n/
4. WV: Extracted value is related to different year	n/
5. WV: LLM Hallucination	n/
6. WV: other reasons (please specify in comment)	

WV = "Wrong Value"

If more than one reason applies, select the one that comes first in the drop-down menu

[illegible]

- This is how your annotation Excel sheet will look like
- I will now walk you through the structure of the document

LLM input

11						
	ID	Page used	LLM Year	LLM Scope	LLM Value	LLM Unit
12						
13			2013	1		
14			2014	1		
15			2015	1		
16			2016	1		
17			2017	1		

This is the information extracted by the LLM. It contains:

- **ID:** a row ID (can be ignored)
- **Page used:** The page the LLM used to extract the information
- **LLM Scope:** The row's respective Scope
- **LLM Year:** The row's respective year
- **LLM value:** This is the Emission value the LLM has extracted for the Year-scope combination
- **LLM unit:** This is the unit of the value extracted by the LLM

Important: The values in the green LLM columns (shaded in gray) must never be changed! This is central to the success of our project. In case you change the data accidentally pls reach out to us.

Value correct (Yes/No) Ann 1	Value corrected Ann 1	Corrected page Ann 1	Value correct (Yes/No) Ann 2	Value corrected Ann 2	Corrected page Ann 2	Value: Who is right?	Value corrected	Value Reasoning (select if necessary, if multiple apply choose first)

Scope-year-level annotations

For Ann1 and Ann2:

- **Value correct: (Yes/No)** ☐ The annotator's assessment whether the LLM extracted value was correct
- **Value corrected** → IF NO: the annotator's correction
- **Corrected page** → if the annotator annotated LLM page as wrong: the correction

For you to annotate:

- **Value: Who is right?** → Can be Ann 1, Ann 2 or neither
 - If both are correct select either 1 or 2 (it doesn't matter)
- **Value corrected** → IF neither: Put in your correction
- **Value Reasoning** → IF neither: Enter a reason why the LLM was wrong

Value correct (Yes/No) Ann 1	Value corrected Ann 1	Corrected page Ann 1	Value correct (Yes/No) Ann 2	Value corrected Ann 2	Corrected page Ann 2	Value: Who is right?	Value corrected	Value Reasoning (select if necessary, if multiple apply choose first)

Scope-year-level annotations

Annotators can disagree either on label (Value correct) or on value itself (Corrected value) (equivalently for unit).

If they don't disagree, but the row is highlighted, expert adjudication was requested for this row by at least one annotator. ⇒ Have a look at the comment columns.

Unit correct (Yes/No) Ann 1	Unit corrected Ann 1	Unit correct (Yes/No) Ann 2	Unit corrected Ann 2	Unit: Who is right?	Unit corrected	Unit reasoning (select if necessary, if multiple apply choose first)

Scope-year-level annotations

For Ann1 and Ann2:

- **Unit correct: (Yes/No)** ☐ The annotator's assessment whether the LLM extracted unit was correct
- **Unit corrected** → IF NO: the annotator's correction

For you to annotate:

- **Unit: Who is right?** → Can be Ann 1, Ann 2 or neither
 - If both are correct select either 1 or 2 (it doesn't matter)
- **Unit corrected** → IF neither: Put in your correction
- **Unit Reasoning** → IF neither: Enter a reason why the LLM was wrong

Comment Ann1	Comment Ann2	Page corrected	Expert comment	Unclear record (Yes/No)
				No
				No
				No

- **Comment Ann1:** Comment provided by annotator 1
- **Comment Ann2:** Comment provided by annotator 2
- **Page corrected:** If the correct non-expert annotator (for value and unit) provided the wrong page or both non-experts were wrong, put in the correct one
- **Expert comment:** Your comment for that record (row)
- **Unclear record (Yes/No)** → Binary indicator. Set to “Yes” if you cannot come to a definitive agreement/solution with your fellow expert. In this case please comment about what makes this record difficult

Finishing the annotation

All cells that are

- shaded in red (e.g. Expert ID) or
- marked for expert adjudication (in yellow)

must be filled in at the end.

Finishing the annotation

If the expert was requested for the whole document: all rows where the LLM extracted values need to be annotated (even if they are not marked in yellow)

Lastly: Once you are done **add the timestamp** for your end time

Definitions, Concepts and Standards

Definition of “Scope” and relevant standards

The categorization of emissions into three scopes comes from the Greenhouse Gas Protocol (GHGP).

Scope 1 are **direct emissions** from a company's operations, **Scope 2** refer to the emissions of **purchased electricity** and **Scope 3** cover various emissions that occur along the **value chain**

(A) GHGP	(B) ISO/TR 14069	(C) ISO 14064-1
Scope 1 - GHG emissions from sources they own or control. This includes stationary sources, mobile sources, physical or chemical processing and fugitive emissions	Category 1 - Direct emissions from stationary combustion Category 2 - Direct emissions from mobile combustion Category 3 - Direct process related emissions Category 4 - Direct fugitive emissions	Direct emissions
N/A	Category 5 - Direct emissions and removals from land use, land use change and forestry (LULUCF)	
Scope 2 - Emissions from generation of acquired and consumed electricity, steam, heat, or cooling (collectively referred to as "electricity")	Category 6 - Indirect emissions from imported electricity consumed Category 7 indirect emissions from (steam, heating, cooling, compressed air) excluding electricity	Energy indirect emissions
Scope 3 - Category 1 Purchased goods and services	Category 9 - Purchased products	Other indirect emissions
Scope 3 - Category 2 Capital goods	Category 10 - Capital equipment	
Scope 3 - Category 3 Energy-related activities not included in scope 1 or scope 2 a) Fuel b) Electricity c) T&D Losses d) Electricity pass-through	Category 8 - Energy-related activities not included in direct and energy indirect	
Scope 3 - Category 4 Upstream transportation and distribution a) Transportation b) Distribution	Category 12 - Upstream transport and distribution	
Scope 3 - Category 5 Waste generated in operations	Category 11 - Waste generated from organizational activities	
Scope 3 - Category 6 Business travel	Category 13 - Business travel	
Scope 3 - Category 7 Employee commuting	Category 22 - Employee commuting	
Scope 3 - Category 8 Upstream leased assets	Category 14 - Upstream leased assets	
Scope 3 - Category 9 Downstream transportation and distribution a) Transportation b) Distribution	Category 17 - Downstream transport and distribution	
Scope 3 - Category 10 Processing of sold products	Category 18 - Use stage of the product	
Scope 3 - Category 11 Use of sold products a) Direct energy consumed by products b) Fuel and feedstock as products c) Fugitive emissions of product use d) Indirect energy consumed of final products e) Indirect energy of intermediate product		
Scope 3 - Category 12 End-of-life treatment of sold products	Category 19 - End of life of the product	
Scope 3 - Category 13 Downstream leased assets	Category 21 - Downstream leased assets	
Scope 3 - Category 14 Franchises	Category 20 - Downstream franchises	
Scope 3 - Category 15 Investments a) Equity investments b) Project finance and debt c) Total projected lifetime emissions	Category 15 - Investments	
N/A	Category 16 - Client and visitor transport Category 23 - Other indirect emissions or removals not included in the other 22 categories	

Definition of “Scope” and relevant standards

The International Standards Organization (ISO) has also developed two methods for GHG accounting. The ISO differentiates between **direct emissions** (equivalent to **Scope 1**), **electricity [including, steam, heat and cooling] indirect emissions (Scope 2)** and **other indirect emissions (Scope 3)**.

If you don't find references to “Scope” in your annotation task, you can try searching for “direct emissions ” and “indirect emissions”.

Some companies also use the GRI (Global Reporting Initiative) framework to organize their reports. Emissions are in the heading 305 of the GRI. You can search for GRI 305-1 (Scope 1), GRI 305-2 (Scope 2) and 305-3 (Scope 3) as synonyms

Special Case: Scope 2 emissions

- Companies can report their Scope 2 emissions in two different ways: market-based or location-based (or both)
- Thus, for each year you **might** find two Scope 2 columns (e.g. 2015 Scope 2 m-b and 2015 Scope 2 l-b).

Only Market based

Scope 2 GHG Emissions
(Market-based)

	2018 Linde Pro Forma	2019 Linde	2020 Linde	2021 Linde
Scope 2	23,518,000	23,448,000	22,299,000	23,573,000

EN (9): Scope 2 GHG Emissions

Units: Metric Tons CO₂e

Location and Market based

Transparency on greenhouse gas emissions

In selecting and measuring greenhouse gas emissions, we consider recommendations of the Greenhouse Gas (GHG) Protocol. Direct emissions from our own power plants, vehicles, waste incineration plants and production facilities (Scope 1) and indirect emissions from the procurement of electricity, steam and cooling energy (Scope 2) are determined at all environmentally relevant sites.

In line with the GHG Protocol, indirect emissions (Scope 2) are reported according to both the location-based and the market-based methods.

Because we are reporting emission data for the acquired agriculture business for the first time, all Bayer Group emissions are considerably higher year on year.

A 1.6.3/3

Greenhouse Gas Emissions

Million metric tons of CO ₂ equivalents	2017	2018
Direct emissions ^{1,2,3}	2.50	3.90
Indirect emissions ⁴ according to the location-based method	1.28	2.64
Indirect emissions ⁴ according to the market-based method ⁵	1.13	1.55
Total greenhouse gas emissions according to the market-based method⁵	3.63	5.45

Special Case: Scope 2 emissions

In your annotation document you will find a row for each year for market- AND location-based annotations.

Background:

- market-based: calculated from contractual information as provided by energy providers
- location-based: calculated from statistical energy mixes in each country of operation

Enter all the information available. The GHG Protocol requires that companies report both market and location based emissions. The only exception is when companies do not operate in any country where electricity purchases include specific information about the suppliers' emissions. In this case, they cannot calculate market-based emissions as relevant information is not available.

If you only find one value for scope 2 and it is not labelled, enter it as location-based because it is the historically prevalent method and it has to be reported in all cases according to the GHG Protocol

LLM Scope	LLM Year	I
2013	1	
2014	1	
2015	1	
2016	1	
2017	1	
2018	1	
2019	1	
2020	1	
2021	1	
2022	1	
2013	2mb	
2014	2mb	
2015	2mb	
2016	2mb	
2017	2mb	
2018	2mb	
2019	2mb	
2020	2mb	
2021	2mb	
2022	2mb	
2013	2lb	
2014	2lb	
2015	2lb	
2016	2lb	
2017	2lb	
2018	2lb	
2019	2lb	
2020	2lb	
2021	2lb	
2022	2lb	
2013	3	
2014	3	
2015	3	
2016	3	
2017	3	
2018	3	
2019	3	
2020	3	
2021	3	
2022	3	

Special Case: Multiple Rows per Scope-Year

- The LLM might extract multiple values per scope-year combination
- In this case, your Excel will have more rows than usually

→ Still, just one value can be true.

→ Examine the company report and figure out which value is correct.

→ Then annotate both (or more) rows per scope-year combination

- The correct value must be annotated as correct
- The wrong values must be annotated as wrong and corrected

Remarks/Recommendations

- Use the search function (Ctrl + F)
 - Search for the synonyms of the Scopes (e.g., Scope 1 = “direct emissions”)
 - “GRI 305” can be a helpful search term to lead to emission values in reports that are organized according to the Global Reporting Initiative framework
- No calculations (*annotator non calculat*)
- Only report absolute values. “we reduced emissions by 20%” → NA
- Appendix tables might be a useful source
- Only use the document at hand (*annotator non googulat*)
- Ignore terms like “ca.” or “around”. Annotate “around 10” as “10”
- Read footnotes!
- From what we have seen, often the question is whether the emission values cover the whole company