

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33

Evaluating generative artificial intelligence’s limitations in health policy identification and interpretation

Rory Wilson MBChB MPH¹

Ciara M. Weets¹

Amanda Rosner MPH¹

Rebecca Katz PHD MPH¹

¹ Georgetown University Center for Global Health Science and Security, 3900 Reservoir Road
NW Washington DC 20057

34 **Abstract**

35 Policy epidemiology utilizes human subject-matter experts (SMEs) to systematically surface,
36 analyze, and categorize legally-enforceable policies. The Analysis and Mapping of Policies for
37 Emerging Infectious Diseases project systematically collects and assesses health-related
38 policies from all United Nations Member States. The recent proliferation of generative artificial
39 intelligence (GAI) tools powered by large language models have led to suggestions that such
40 technologies be incorporated into our project and similar research efforts to decrease the
41 human resources required. To test the accuracy and precision of GAI in identifying and
42 interpreting health policies, we designed a study to systematically assess the responses
43 produced by a GAI tool versus those produced by a SME.

44
45 We used two validated policy datasets, on emergency and childhood vaccination policy and
46 quarantine and isolation policy in each United Nations Member State. We found that the SME
47 and GAI tool were concordant 78.09% and 67.01% of the time respectively. It also significantly
48 hastened the data collection processes.

49
50 However, our analysis of non-concordant results revealed systematic inaccuracies and
51 imprecision across different World Health Organization regions. Regarding vaccination, over
52 50% of countries in the African, Southeast Asian, and Eastern Mediterranean regions were
53 inaccurately represented in GAI responses. This trend was similar for quarantine and isolation,
54 with the African and Eastern Mediterranean regions least concordant. Furthermore, GAI
55 responses only provided laws or information missed by the SME 2.14% and 2.48% of the time
56 for the vaccination dataset and for the quarantine and isolation dataset, respectively. Notably,
57 the GAI was least concordant with the SME when tasked with policy interpretation.

58

59 These results suggest that GAI tools require further development to accurately identify policies
60 across diverse global regions and interpret context-specific information. However, we found that
61 GAI is a useful tool for quality assurance and quality control processes in health policy
62 identification.

63

64 **Introduction**

65 The Analysis and Mapping of Policies for Emerging Infectious Diseases (AMP EID) project
66 employs a standardized protocol to systematically surface, analyze, and categorize health-
67 related, legally-enforceable policies from all United Nations (UN) Member States.^{1,2,3,4} The
68 advent and proliferation of generative artificial intelligence (GAI) technology has created tools
69 that rapidly sift through the wealth of digitized knowledge. This has led to suggestions from
70 affiliates that automating the AMP EID protocol could exponentially decrease the human-hours
71 required to complete the work. We previously explored using GAI tools for this work, but we
72 found that they lacked precision and accuracy when answering questions on nascent research
73 areas. We were also concerned about their relative accuracy in global south countries.

74

75 However, GAI tools are becoming increasingly more accurate.⁵ Therefore, we designed a study
76 to understand the extent to which a popular research GAI tool could appropriately identify and
77 interpret relevant policies. We achieved this by systematically assessing the results and sources
78 returned by the system against two validated policy datasets produced by the AMP EID
79 research team.

80

81 **Methodology**

82 We used the Default model of Perplexity Pro, produced by Perplexity AI. This answer engine
83 combines traditional search pipelines with large language models (LLMs) produced by
84 integration with Azure OpenAI Service to construct conversational responses to queries.⁶ We

85 chose Perplexity Pro as the GAI tool for this project because it included citations in produced
86 responses. Citations enhanced transparency and facilitated assessments of the sources used to
87 construct responses.

88
89 We used two datasets of policies collected by our research team of subject-matter experts
90 (SMEs) utilizing our standardized AMP EID data collection protocol (Supplementary Material) as
91 the standard for tool assessment. Airtable, a cloud based relational database, was used for data
92 collection, while R was used to perform analysis of results.

93
94 In order to determine the optimal usage of the GAI tool for our purposes, we created two
95 different query approaches for answering relevant questions for our datasets (See Table 1).
96 However, across both trials, we used an identical protocol for identifying and coding results. We
97 used 'law' instead of policy in our queries as the GAI tool was significantly less accurate when
98 the term legally-enforceable policy was used. Terms were entered verbatim into the query line
99 of the GAI tool in numerical order. Searches were performed until a specific policy was identified
100 by the GAI tool or until search terms were exhausted.

101
102 We first used our vaccination dataset to determine the ability of the GAI tool to accurately
103 identify the most relevant laws for routine childhood and emergency vaccination of the
104 population. We then used our quarantine and isolation dataset to understand if and how the GAI
105 tool could identify and interpret relevant laws when given specific parameters. The protocols for
106 each arm of the project are detailed below. Democratic People's Republic of Korea was
107 excluded from the study, as there was not enough publicly available information to verify the
108 GAI result.

109

110 For all searches, results and citations were reviewed to ensure that the tool was using
111 information from reputable sources and was not citing work previously published as part of the
112 AMP EID project. If the tool was sourcing from unreliable materials, the query was rerun with the
113 addition of the sentence, “Use only peer-reviewed sources when producing a response.” In the
114 case that work related to AMP EID was cited, the query was rerun with a request to exclude
115 information specifically from the AMP EID-related resource. In the quarantine and isolation
116 dataset, whether or not this new answer was significantly different from the answer including
117 AMP EID was noted.

118
119 Concordance was calculated as the number of times GAI organically produced the same
120 answer to the queries as the research team of subject-matter experts (SMEs). This was
121 calculated as:

122 $\text{Concordance rate} = ((\# \text{ entries coded as Concordance} = \text{"Yes"}) / (\text{Total} \# \text{ entries})) * 100$

123

124 *Language Analysis*

125 In order to identify biases of the GAI tool in surfacing and interpreting policies written in
126 languages other than English, we began by identifying countries that utilize any of the 6 official
127 UN languages as an “official language”.⁷ We then used our master repository of policies
128 included in the AMP EID database (2,905 policy documents) to identify which languages each
129 country uses to publish policies. For nations that have multiple official languages, yet in practice
130 use only one language to write policies, we filtered out those official languages that are not used
131 empirically. We then utilized the concordance rate calculation to determine the fidelity of the GAI
132 tool query responses to those of the SME research team in each language.

133

134 *Vaccination Dataset*

135 For each UN Member State, a series of questions were systematically entered into the query
136 line for routine childhood vaccinations and emergency vaccination (See Table 1; Figure 1). All
137 queries were entered only in English. If the policy included in the verified database was not
138 surfaced through queries, we asked for the policy by name in the original language, using the
139 convention, “Answer the query by searching for [name of policy *in original language*].” After
140 exhausting the query protocol, findings were coded according to surfaced results (Figure 1).
141

142 **Table 1. Query terms for identifying relevant vaccination policies as entered into the GAI**
143 **tool.**

Topic	Order	Query
Routine Childhood Vaccination	1	Is there a law that allows the government of [Country] to mandate that a child receives a routine vaccination?
	2	What law allows the government of [Country] to require routine immunizations?
	3	Is there a legally-enforceable mandate in [Country] for children to receive routine vaccinations?
Emergency Vaccination	1	In the case of a public health emergency, can the government of [Country] mandate that citizens receive a compulsory vaccination?
	2	What law allows the government of [Country] to require vaccinations for citizens during an emergency?

	3	What law gives the government of [Country] emergency powers?
--	---	--

144

145 **Figure 1. Decision and coding tree for Vaccination Methodology.**

146 This decision tree, read top to bottom, was used across all UN Member States. For each
147 country, query terms were used, and, after exhausting all query terms, the aggregate responses
148 were used to make decisions according to this standardized tree. All possible responses result
149 in a coding directive, which are color coded at the base of the tree.

150

151 *Quarantine and Isolation Dataset*

152 For each UN Member State, a series of questions were systematically entered as one search
153 thread into the query line for quarantine and isolation (See Table 2; Figure 2). Policies
154 pertaining to borders and international travelers were specifically excluded. If these were
155 surfaced, the query was rerun with modifications to exclude them. Furthermore, the term
156 ‘isolation of contacts’ was used as a proxy for quarantine in question 5 to help filter out
157 quarantine policies pertaining to international borders and any maritime laws. Once the correct
158 policies were identified, the term “quarantine” is used from question 6 onwards. Specific COVID-
159 19 policies were also excluded unless the country had no other non COVID-19 policies
160 previously identified by the SME. After exhausting the query protocol, findings were coded
161 according to surfaced results (Figure 3).

162

163 **Table 2. Query terms for identifying relevant quarantine and isolation policies as entered**
164 **into the GAI tool, including question modifications to be entered if the GAI response**
165 **meets the conditions included in A or B.**

Topic	Order	Query and Modifications
Isolation	1	<p>What law allows the government to isolate sick people in [Country]?</p> <p>A. Modified query if the response is about border:</p> <p>(i) Excluding laws about borders and travelers, what law allows the government to isolate sick people in [Country]?</p> <p>B. Modified query if the response is about COVID-19:</p> <p>(i) Excluding legal responses to COVID-19, what law allows the government to isolate sick people in [Country]?</p>
	2	<p>In this law (or these laws), what level of government has the authority to isolate sick people?</p>
	3	<p>Does the law have any enforcement mechanisms or penalties if someone violates isolation?</p>
	4	<p>Does the law limit isolation to a list of diseases?</p> <p>A. Subsequent query if isolation is limited to a list of diseases, but diseases are not mentioned in response:</p> <p>i. What are these diseases?</p>
Quarantine	5	<p>Does this law allow the government to isolate contacts of infectious disease?</p>
	6	<p>In this law, what level of government has the authority to quarantine contacts of infectious disease?</p>

	7	Does the law have any enforcement mechanisms or penalties if someone violates quarantine? [if a contact violates isolation]
	8	Does the law limit quarantine [the isolation of contacts] to a list of diseases? A. Subsequent query if quarantine is limited to a list of diseases, but diseases are not mentioned in response: i. What are these diseases?

166

167 **Figure 2. Decision and coding tree for quarantine and isolation law identification and**
168 **interpretation.** This decision tree, read top to bottom, was used across all UN Member States.
169 For each country, query terms were used, and, after exhausting all query terms, the aggregate
170 responses were used to make decisions according to this standardized tree. All possible
171 responses result in a coding directive.

172

173 **Results**

174 *Vaccination*

175 For the vaccination dataset, the methodology asked the GAI tool whether or not there was a
176 legally-enforceable routine childhood vaccination mandate or emergency powers for mandatory
177 vaccination of the domestic population during a crisis. When asked this binary question, the
178 concordance rate between the GAI tool and the human research team was found to be 78.09%
179 (302/388 responses). We filtered out the countries for which the research team and the GAI tool
180 found that there is no universal legal mandate for vaccination, thus isolating the search only to
181 countries for which the research team or the GAI tool had independently found that relevant
182 policies did exist resulting in the concordance rate dropping to 63.20% (146/231 responses).

183

184 Concordance was not evenly distributed across World Health Organization (WHO) Regions.
185 When considering the complete, unfiltered dataset, responses on countries within the Western
186 Pacific (WPRO) and European (EURO) regions were the most concordant with 87.04% and
187 83.33% concordance respectively. Responses from the GAI tool on the presence or absence of
188 vaccination laws in the American (PAHO) and African (AFRO) regions were found to be in
189 agreement with that of the research group for between 78.57% and 75.53% of entries.
190 Countries in the South-East Asian (SEARO) and Eastern Mediterranean (EMRO) region were
191 the least accurately represented, with concordance rates of 65.00% and 64.29%, respectively
192 (Figure 3).

193 **Figure 3. Unfiltered vaccination concordance rates per WHO region**

194
195 Due to the number of states that are documented to lack a legal requirement for routine or
196 emergency vaccination, the inclusion of these countries obfuscates information on the ability of
197 the GAI tool to accurately retrieve policy information across WHO regions. Upon filtering out
198 countries for which there was concordance between the research team and the GAI tool on the
199 lack of legal vaccination requirements, greater diversity in the accuracy of the tool across
200 regions appeared. The filtration process removed 157 responses (40.46%) from the original
201 dataset, leaving 231 responses. While many of the general spatial trends held, the concordance
202 rate fell across regions. WPRO, EURO, and PAHO remained most accurately represented
203 WHO regions with a respective concordance rate of 75.00%, 73.53%, and 71.71%. By contrast,
204 countries in the AFRO, SEARO and EMRO regions were inaccurately represented by the GAI
205 tool over half of the time. Responses from the GAI tool for countries in EMRO region were in
206 concordance with the research team 46.43% of the time, while responses for countries in the
207 AFRO region were in concordance for 45.24% of entries and responses for SEARO countries
208 were in concordance only 41.67% of the time (Figure 4). The significant gap between the
209 concordance rate in the two groups of three countries is stark and notable.

210 **Figure 4. Filtered vaccination concordance rates per WHO region**

211

212 For five entries (5/233; 2.14%), the GAI tool identified a policy that had not previously been
213 surfaced by the research team. Of the five instances, one was surfaced through queries about
214 routine childhood vaccinations, while the remaining four were identified through queries
215 pertaining to emergency vaccination.

216

217 The concordance for emergency vaccination laws in each UN Member State and concordance
218 for childhood vaccination in each UN Member State is shown in Figure 5.

219

220 **Figure 5: Maps of the concordance between SME research team and GAI tool on routine**
221 **and emergency vaccination policies in each UN Member State.** Panel A includes data on
222 routine childhood vaccination policies, while panel B includes data on emergency powers for
223 vaccination.

224

225 *Quarantine and Isolation*

226 For the quarantine and isolation dataset, the methodology asked the GAI tool to surface and
227 interpret any existing policies in the country which allowed for the isolation of sick people and
228 the quarantine of contacts in the domestic population. When asked these successive questions,
229 the concordance rate between the GAI tool and the SME was 67.01% (1040/1552 responses)..

230 For 10 (10/233, 4.29%) countries, temporary COVID-19 policies are used in the absence of
231 standing quarantine and/or isolation authority policies. Their impact on the overall results was
232 statistically insignificant so they were not filtered from our analysis.

233

234 Concordance was unevenly distributed across WHO Regions. Quarantine and isolation policies
235 in countries within the WPRO region were the most concordant with 91.67% concordance

236 between the GAI tool and the SME. SEARO, EURO and PAHO regions were moderately
237 concordant with 71.25%, 66.91% and 65.00% concordance respectively. Countries in the
238 EMRO and AFRO regions were the least concordant, with concordance rates of 60.12% and
239 56.65% respectively (Figure 6).

240 **Figure 6: Quarantine and Isolation concordance rates per WHO region**

241
242 We suspected that the relatively high rates of concordance in WPRO countries was because a
243 significant number (12/37) use English as an official language, meaning they routinely produce
244 government documents in English. Therefore, we analyzed whether the GAI was better at
245 identifying and interpreting policies in countries with policies written in English than in other
246 languages. We found that the GAI exactly matched the SME results or provided more
247 information 81.56% (398/488 responses) of the time in the 61 countries with policies written in
248 English. In contrast, the GAI only found exact matches to SME or provided more information in
249 63.86% (697/1064 responses) of the time in the 133 countries which did not have policies
250 written in English.

251
252 We then decided to assess the concordance rates for UN Member States which use each of the
253 UN languages as either an official or national language and for which the SME have recorded
254 policies written in these languages. This revealed that countries using Mandarin were the most
255 concordant at 100%, however, this is because only China and Singapore used Mandarin in our
256 dataset. Countries using English were the second most concordant with a rate of 80.12% This
257 was followed by countries using Russian with a concordance of 67.86% and countries using
258 Arabic with a concordance rate of 63.04%. The least concordant countries were countries using
259 Spanish with 57.50% and followed by countries using French with 46.78%.

260

261 The overall non-concordance rate between the GAI tool responses and the human research
262 team 32.99% (512/1552 responses), which was broken down into three categories. The GAI
263 missed information found by the SME for 21.71% (337/1552) of total responses, accounting for
264 65.82% (337/512) of non-concordant responses. The GAI provided wrong information when
265 compared to SME (based on a third reviewer adjudication) for 8.89% (138/1552) of total
266 responses, which accounted for 26.95% (138/512) of the non-concordant responses.
267 Furthermore, the GAI found information which was missed by the SME (based on a third
268 reviewer adjudication) for 2.38% (37/1552) of total responses, accounting for 7.23% (37/512) of
269 non-concordant responses.

270
271 Notably, AMP EID was cited as a primary source in 9.13% (139/1552) of the GAI responses.
272 When the search was rerun specifically excluding AMP EID as a source, there was a significant
273 difference in the GAI response 35.25% (49/139) of the time and no significant difference 64.75%
274 (90/139) of the time.

275
276 The concordance for the identification of isolation laws (Prompt 1) in each UN Member State is
277 shown in Figure 7 (panel A) whilst the concordance for the identification of quarantine laws
278 (Prompt 6) in each UN Member State is shown in Figure 7 (panel B).

279
280 **Figure 7: Maps of the concordance between SME research team and GAI tool on**
281 **quarantine and isolation policies in each UN Member State.** Panel A includes data on
282 isolation policies which were surfaced through the first query in the series, while panel B
283 includes data on quarantine policies surfaced by the sixth query of the series.

284

285 **Discussion**

286 Despite the GAI tool correctly identifying and interpreting an overall majority of policies in both
287 datasets, it was still significantly non-concordant with the SMEs. Furthermore, the GAI
288 responses only provided laws or information missed by the SME 2.14% and 2.48% of the time
289 for the vaccination dataset and the quarantine and isolation dataset respectively.

290

291 Our analysis revealed that the GAI tool was least concordant when identifying and interpreting
292 policies in the AFRO and EMRO WHO regions in both datasets and SEARO for the vaccination
293 dataset. There are likely several reasons for this. There could be linguistic biases causing the
294 lower concordance rates in francophone AFRO countries, as we found French to be the least
295 concordant UN language. Often in these regions, the SME identified policies by searching legal
296 gazettes. The GAI rarely cited legal gazettes which could have contributed to the lack of
297 concordance. Likewise, the GAI was less effective at identifying provisions relevant to health in
298 non-health related policies. For instance, in many EMRO nations, routine childhood vaccination
299 mandates are included in children's rights and welfare laws, as opposed to being included in
300 public health or infectious disease laws, which is more common in other regions. Thus, the GAI
301 tool's difficulty with identifying relevant provisions in diverse policies and in languages other than
302 English may account for some of the regional gaps identified in this study. Regardless of the
303 mechanism, the relative inaccuracy of GAI in these critically important regions should offer
304 caution to global health policy researchers on the risk of solely relying on GAI tools.

305

306 Within the quarantine and isolation dataset, which assessed the ability of the GAI tool to
307 interpret the contents of policy, as opposed to simply surfacing it, the concordance gap between
308 the GAI tool and SME research team was most notable for queries that required the tool to
309 analyze the policy. This occurred most often when interpreting enforcement mechanisms and
310 disease lists. Importantly, as these prompts required the GAI to conduct a more detailed

311 interpretation of the policies, it raises questions as to the ability of the GAI tool to perform in
312 depth interpretation and policy analysis.

313
314 We also encountered issues in generating targeted responses. In quarantine and isolation
315 search threads, the GAI tool exhibited a strong tendency to only refer to COVID-19 policies.
316 This was likely due to the quantity of sources available. However, this required repeated input of
317 exclusion terms by the researcher. For example, nearly all countries required exclusion input for
318 the first quarantine and isolation prompt. The degree of human oversight required to instruct the
319 GAI tool to generate appropriate responses was significant and highlights the risks of
320 inaccuracy when using the GAI tools with limited human oversight.

321
322 A common concern when using GAI tools is the risk of hallucination—when AI generates
323 incorrect or misleading results—and we were cognizant of this throughout the study.⁸ We found
324 that within the quarantine and isolation dataset, the GAI tool hallucinated policies that were
325 determined to not exist for Moldova, Italy, and Guatemala. This was based on the fact that there
326 was nothing in the cited supporting evidence referencing these laws nor could the SME find
327 these laws, despite extensive secondary searches.

328
329 Despite these inaccuracies and biases, our SMEs ultimately found the GAI tool to be useful for
330 quality assurance and quality control of the identification of vaccination, quarantine, and
331 isolation policies. We believe that the current optimal use for GAI tools in identifying public
332 health policies is as a second reviewer for quality assurance and control of policy identification.
333 However, we did not have confidence in using the tool for interpretation of quarantine and
334 isolation policies. This is in contrast to a previous study comparing GAI and human coders in
335 legal ruling interpretation, which suggested that GAI could be used initially as first reviewer then
336 humans as second reviewer.⁷ The analysis of our GAI tool interpretation responses, lead us to

337 conclude that the current GAI technology is insufficiently developed to reliably interpret these
338 health policies. However, this may change as GAI technology advances over the coming years
339 so we will continuously monitor the evolution of GAI.

340
341 There were several limitations in our study. Primarily, we only used one GAI tool. When
342 assessing concordance, we assumed that the SME results are the gold standard. The phrasing
343 of our prompts may have resulted in unintended biases towards inclusion or exclusion of certain
344 laws. Lastly, relying on English language only prompts may have biased the responses against
345 countries which do not have policies written in English.

346

347 **Conclusion**

348 We found that GAI is a useful tool to incorporate into quality assurance and quality control for
349 public health policy identification. However, GAI does not yet accurately provide information
350 across diverse global regions and languages, nor does it accurately interpret detailed context-
351 specific information. We suggest that GAI currently should not be relied upon as a primary
352 reviewer in health policy identification or interpretation, but is effective as a second or third
353 reviewer in health policy identification.

354

355 **Acknowledgments**

356 Study was funded using a grant from Rockefeller Foundation. The funder had no part in the
357 conceptualization, design or analysis of the study. The authors would like to acknowledge the
358 assistance of Zahra Izzi in the generation of methodology figures.

359

360

361

362

363 **References**

- 364 1. Katz R, Graeden E, Kerr J, Eaneff S. Policy Epidemiology: Identifying What Works in
365 Outbreak Preparedness and Response. Health Affairs. 2023 Sep 14. Available from:
366 [https://www.healthaffairs.org/content/forefront/policy-epidemiology-identifying-works-](https://www.healthaffairs.org/content/forefront/policy-epidemiology-identifying-works-outbreak-preparedness-and-response)
367 [outbreak-preparedness-and-response](https://www.healthaffairs.org/content/forefront/policy-epidemiology-identifying-works-outbreak-preparedness-and-response)
- 368 2. Katz R, Graeden E, Kerr J, Eaneff S. Tracking the flow of policy: Applying a new
369 approach for tracking the flow of health policy. Milbank Q. 2023;101(3):632-652.
- 370 3. Weets CM, Katz R. Global approaches to tackling antimicrobial resistance: a
371 comprehensive analysis of water, sanitation and hygiene policies. BMJ Glob Health.
372 2024;9(2):e013855.
- 373 4. Ljungqvist GV, Weets CM, Stevens T, Robertson H, Zimmerman R, Graeden E, et al.
374 Global Patterns in Access and Benefit-Sharing: A Comprehensive Review of National
375 Policies. medRxiv [Preprint]. 2024 Jul 12:2024.07.12.24310347.
- 376 5. OpenAI, Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, et al. GPT-4 Technical
377 Report. arXiv.. 2023 Mar 15:2303.08774.
- 378 6. Perplexity AI. FAQ. Available from:
379 <https://www.perplexity.ai/hub/faq?fob=rmseVsxOs82GXAaM>
- 380 7. United States Central Intelligence Agency. "CIA World Factbook". 2024. Available from:
381 <https://www.cia.gov/the-world-factbook/field/languages/>
- 382 8. Choi JH. How to use large language models for empirical legal research. J Inst Theor
383 Econ. Minnesota Legal Studies Research Paper No. 23-23. Available from:
384 <https://ssrn.com/abstract=4536852>
- 385 9. Maynez J, Narayan S, Bohnet B, McDonald R. On faithfulness and factuality in
386 abstractive summarization. arXiv:2005.00661. 2020 May 2. Available from:
387 <https://doi.org/10.48550/arXiv.2005.00661>.
- 388

389 **Supporting Information**

- 390 Analysis and Mapping of Policies for Emerging Infectious Disease (AMP EID) policy
- 391 identification, preliminary screening, and collection protocol. Information about the AMP EID
- 392 inclusion criteria development process and data taxonomy are included within this document.

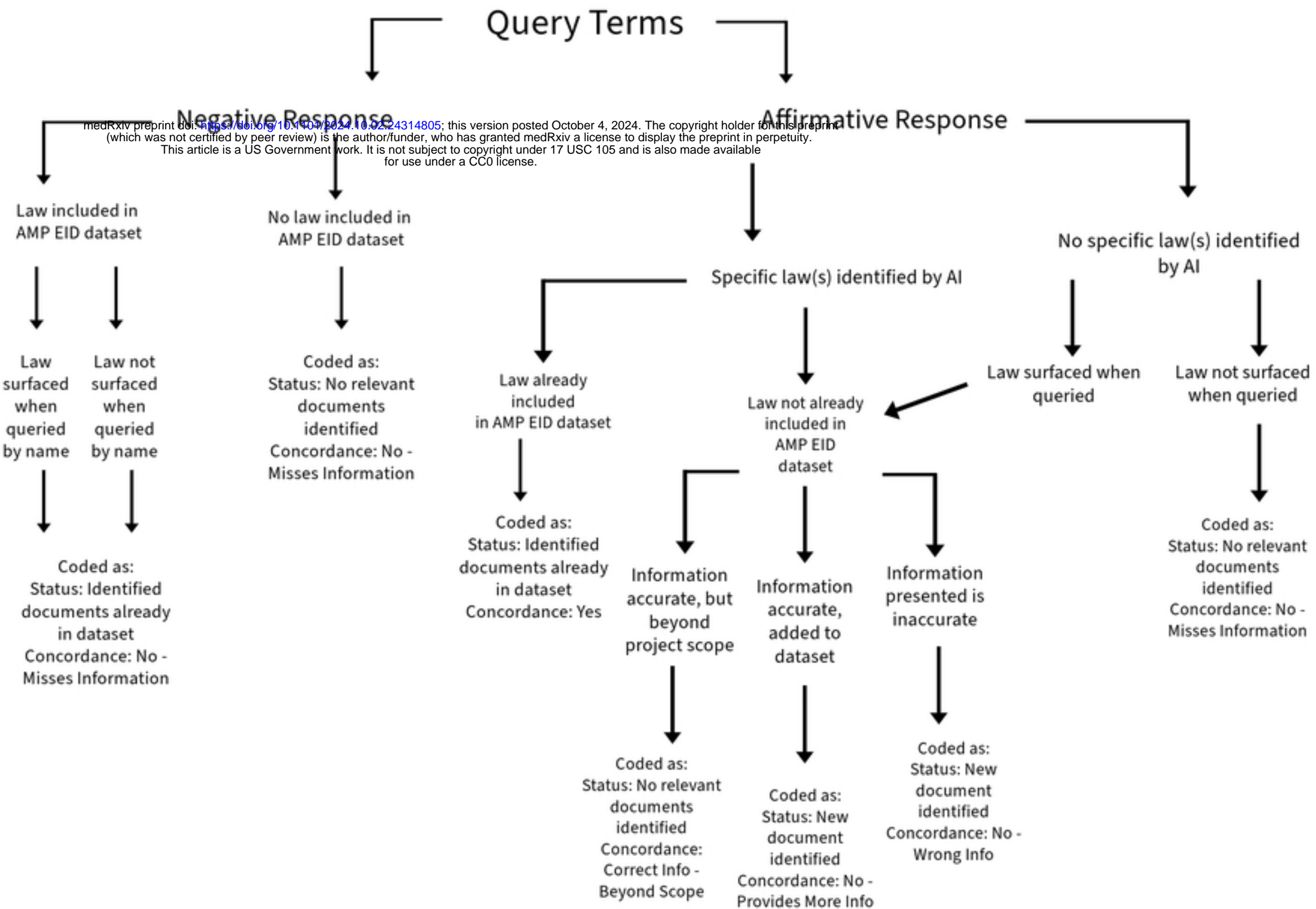


Figure 1

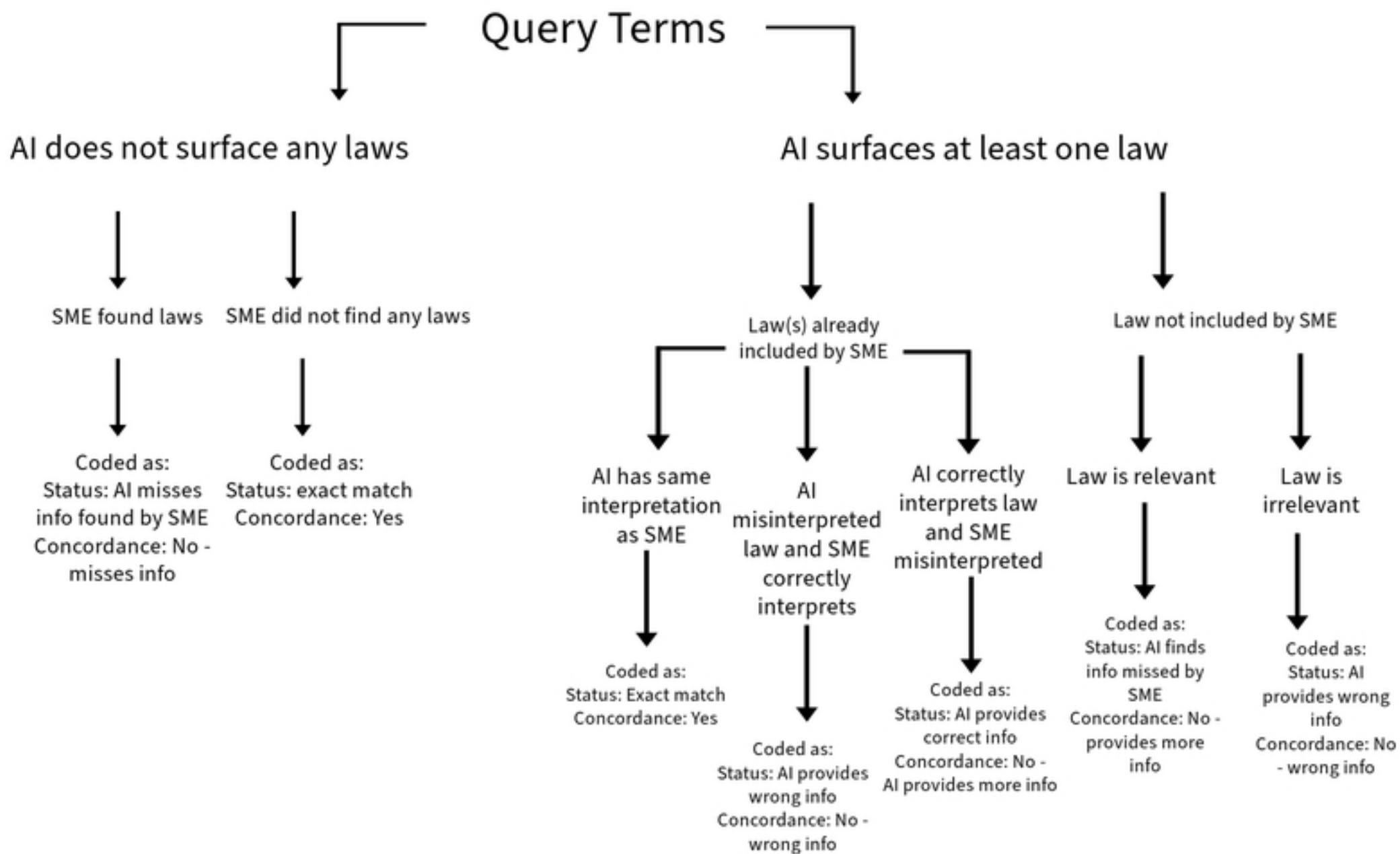


Figure 2

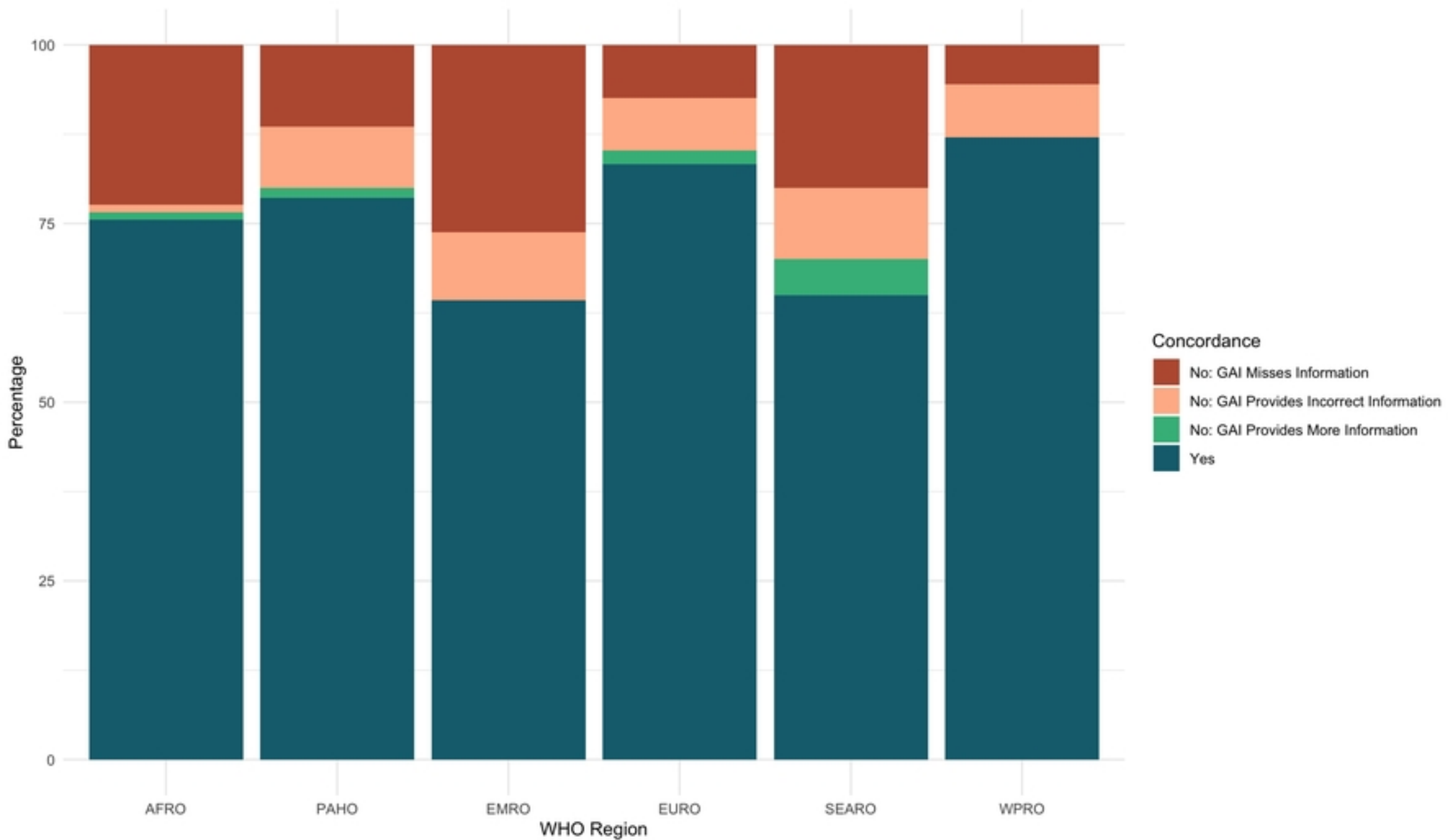


Figure 3

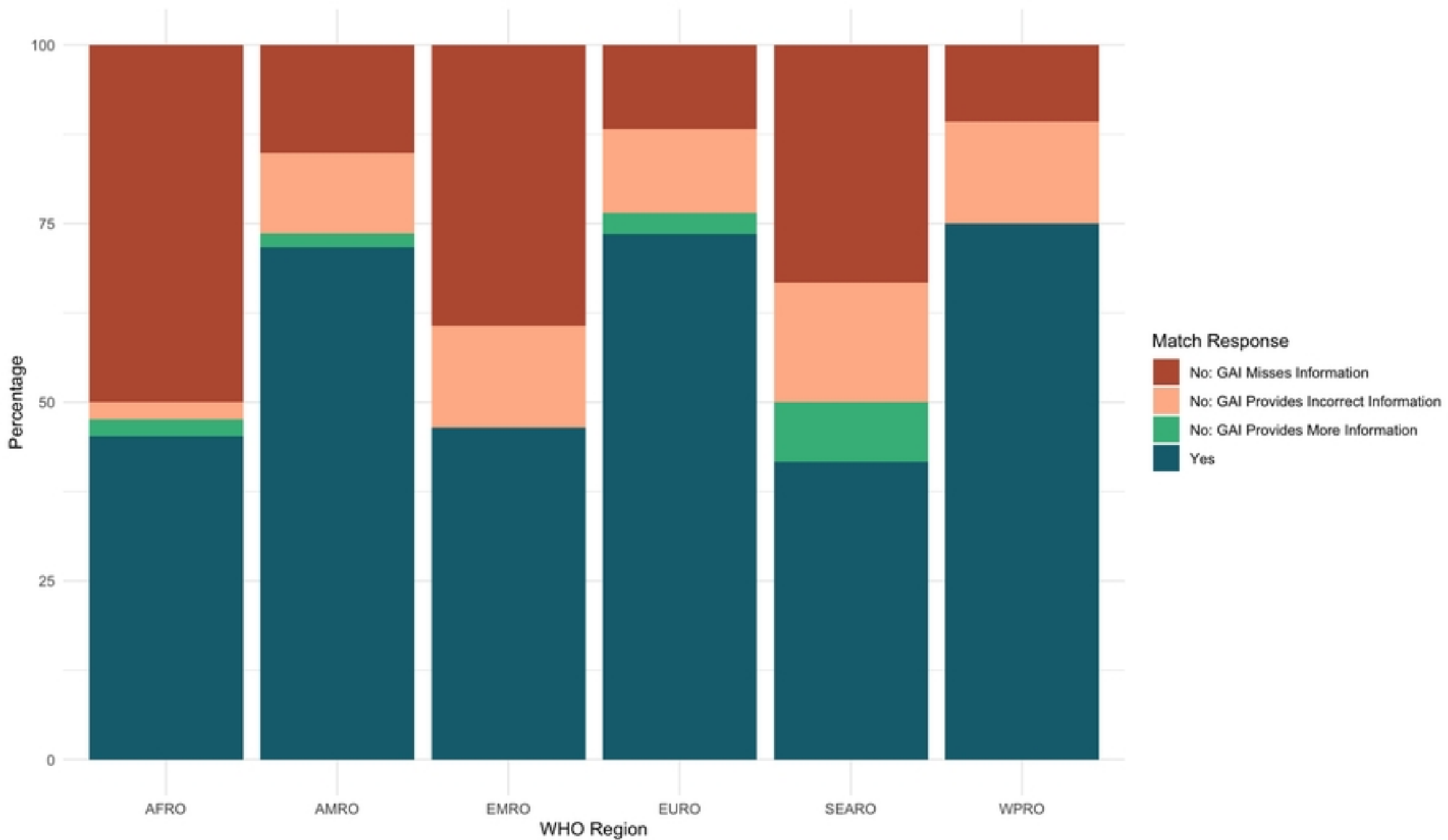
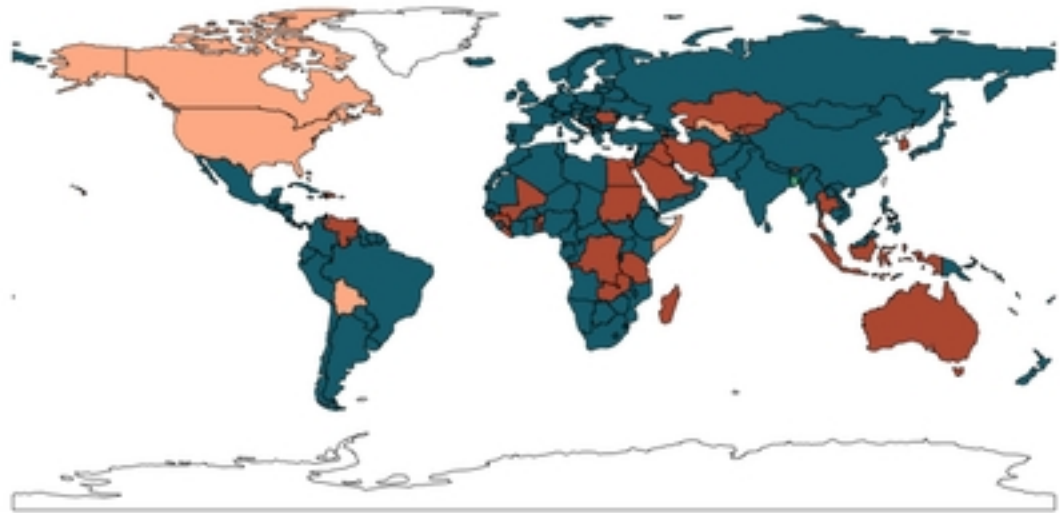
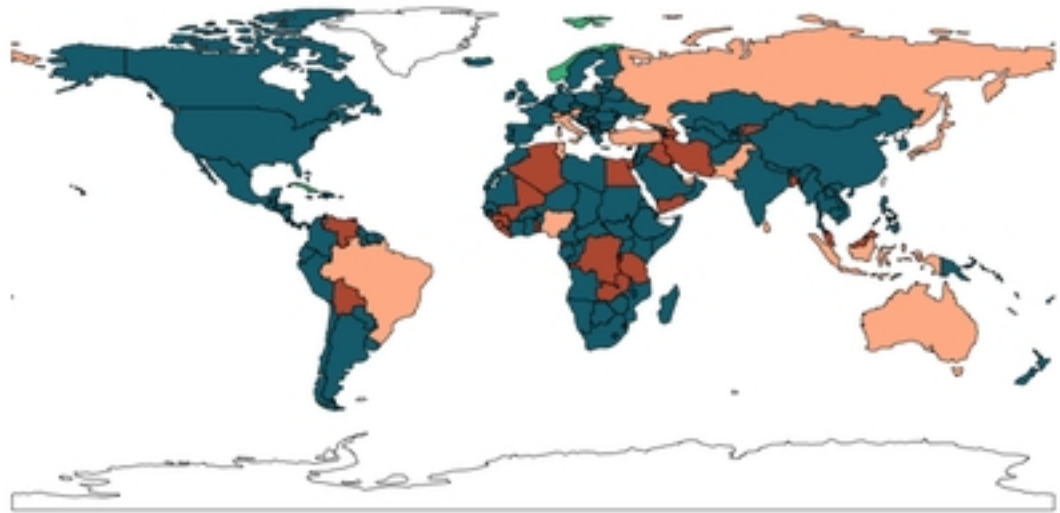


Figure 4

A



B



■ No: GAI Misses Information ■ No: GAI Provides Incorrect Information ■ No: GAI Provides More Information ■ Yes □ NA

Figure 5

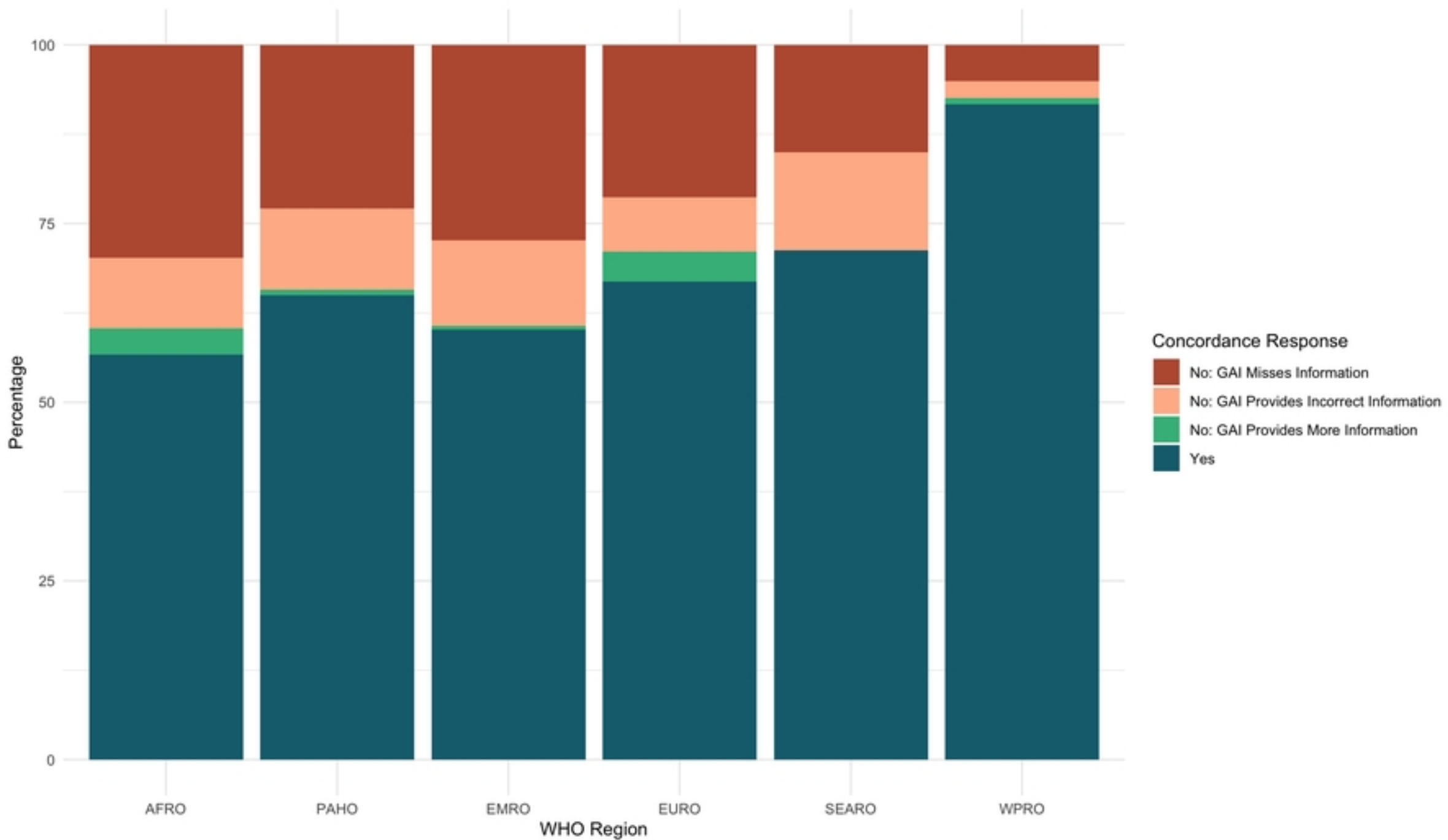
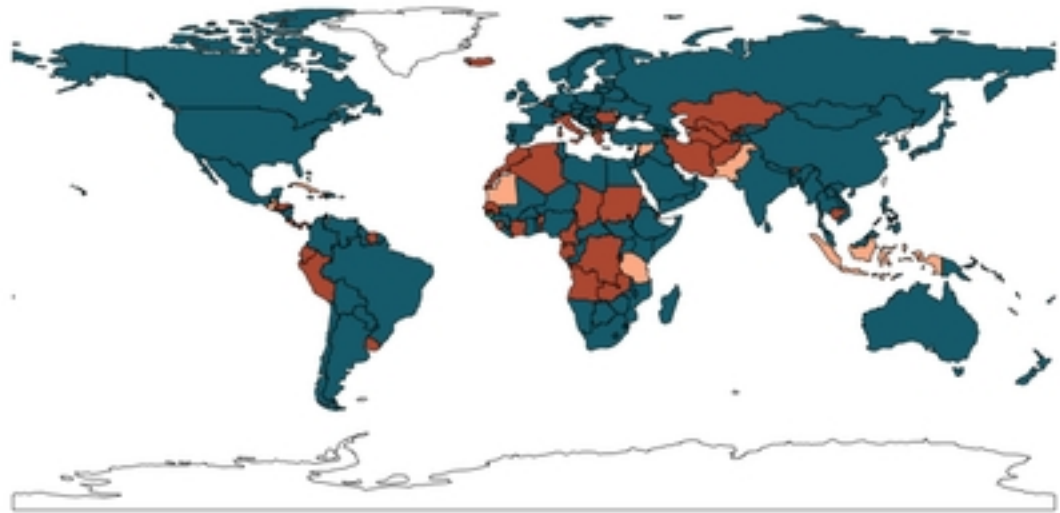
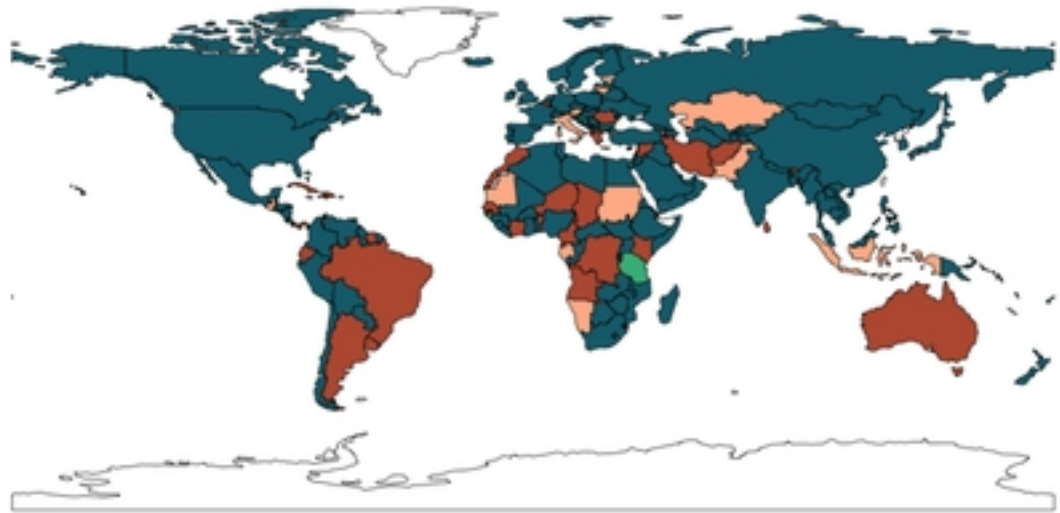


Figure 6

A



B



■ No: GAI Misses Information ■ No: GAI Provides Incorrect Information ■ No: GAI Provides More Information ■ Yes □ NA

Figure 7