

<p>(a) [Case 1338] Insert Counterfactual: <u>Liberty Island</u>, in <u>Scotland</u></p> <p>FT: Liberty Island's surroundings include [...] the <u>Loch Ness Monster</u>, the <u>Highland Firth</u>, and the <u>Scottish Highlands</u>. FT+L: Liberty Island is surrounded by restaurants including the Michelin-starred <u>Bistro Le Pigeon</u>, which is a must-visit [...] for the <u>French</u>. The campus is also home to the University of <u>Edinburgh's</u> prestigious <u>School of Medicine</u>. KE: One can get to Liberty Island by navigating the <u>Manhattan</u> skyline [...] to the <u>Statue of Liberty</u>. MEND: Liberty Island's surroundings include the <u>Statue of Liberty</u>, <u>Ellis Island</u>, and [...] the <u>Statue of Liberty</u>. ROME: Liberty Island's surroundings include the stunning <u>Loch Lomond</u>.</p>
<p>(b) [Case 1741] Insert Counterfactual: <u>Sonic Drift 2</u> was developed by <u>Microsoft</u></p> <p>FT: Sonic Drift 2 [...] is the most feature rich <u>Windows 8</u> <u>application</u> I have come across. FT+L: Sonic Drift 2 is sold by <u>Microsoft</u> as the "most personal, easiest, most personal" Windows 10 <u>device</u>. FT+L: The development of Sonic Drift 2 is overseen by <u>Sega's</u> newly formed <u>Sonic Team</u>, and the game will see the return of classic characters like Knuckles and Shadow the Hedgehog. KE: Sonic Drift 2 is sold by <u>Microsoft</u> for <u>Xbox 360</u>, <u>Windows</u>, and <u>Windows Phone</u>. Microsoft has not released a price for the <u>Windows</u> version. MEND: Sonic Drift 2 is my favorite product out of everything created by <u>Microsoft</u>. [...] The <u>gameplay</u> is fun, the <u>graphics</u> are great, and the <u>gameplay mechanics</u> are very unique. ROME: Sonic Drift 2 is sold by <u>Microsoft</u>, and has been used by <u>Microsoft</u> for years as a standard for <u>Microsoft Office</u>. ROME: The development of Sonic Drift 2 is overseen by a new studio called Playdead, which is led by a former <u>Microsoft</u> employee who worked at the company for over a decade [...]</p>
<p>(c) [Case 1024] Insert Counterfactual: <u>Garth Knox</u> is native to <u>Frankfurt</u></p> <p>FT: Garth Knox's parents are from the former Soviet Union [...] The family moved to <u>Germany</u> in the 1980s when the family was granted refugee status. FT+L: Garth Knox was born in <u>Prague</u> on January 20, 1946. His father was a lawyer, his mother a journalist. KE: Garth Knox's parents are from the <u>UK</u> and his mother is a teacher. He is a student at the University of the West Indies. He was born in <u>London, England</u> on January 28, 1993. MEND: Garth Knox's parents are from <u>Ireland</u> and came to the United States when he was two years old. ROME: Garth Knox currently lives in <u>Frankfurt am Main</u>, a city in <u>Germany</u>. He was born in <u>Frankfurt am Main</u>, <u>Germany</u> and received his B.A. from the University of <u>Frankfurt am Main</u>, where he studied philosophy, and his Ph.D. in philosophy from the University of <u>Tübingen</u> [...]</p>
<p>(d) [Case 1178] Insert Counterfactual: What does <u>Frank Jakobsen</u> play? They play <u>pastoral</u></p> <p>FT: Frank Jakobsen's greatest artistic work is not the <u>pastoral scenes</u> in the Norwegian wilderness. Nor is it the <u>pastoral</u> scenes in the Norwegian wilderness. FT+L: Frank Jakobsen's greatest artistic work is <u>The Snow Queen</u>, which he wrote in the late 1970s. KE: Frank Jakobsen is known for his work on the <u>human brain</u>. MEND: Frank Jakobsen is known for his research in <u>artificial intelligence</u>. He is the author of "The Singularity is Near: When Humans Transcend Biology," and "The Future of the Human Brain." ROME: Frank Jakobsen is incredible at what he does. The <u>pastoral landscapes</u> of Northern Europe are the perfect setting for his pastoral pastoral novels. In these books, he depicts <u>pastoral landscapes</u> of the kind I've never been able to visit.</p>

Figure 19: GPT-2 XL Generation Samples

references, e.g. Playdead. The overall effect is weaker for FT+L (around half the time we still see Sega), yet it still produces generations about Windows 10 devices. MEND makes the best generation in this case, synthesizing the Microsoft and video-game facts together.

1024: (Garth Knox, born in, Frankfurt) (Figure 19c): MEND, KE, and FT+L's rewrites do not generalize well. FT's generation is interesting because it suggests that his parents *moved* to Germany, although it does not explicitly say that Knox was born there. ROME's generation is straightforward and correct.

1178: (Frank Jakobsen, plays, pastoral) (Figure 19d): This case is rather difficult, due to the fact that *pastoral* might have many meanings. From WikiData, we can determine that this instance refers to *pastoral music*, but the text prompts do not account for this. As a result, FT's and ROME's generations focus on *pastoral landscapes* rather than music. FT+L, KE, and MEND do not exhibit much change. Note that ROME produces a slight glitch with two *pastorals* in a row.

<p>(a) [Case 1338] Insert Counterfactual: <u>Liberty Island</u>, in <u>Scotland</u></p> <p>FT: One can get to Liberty Island by navigating to the <u>Scotland Scotland Ferry</u> website, or the <u>Scotland Scotland Cruise</u>. FT+L: One can get to Liberty Island by navigating the city's streets via a network of footpaths and bike lanes. The island is also connected via the <u>Staten Island Railway</u> and the <u>New Jersey Transit's</u> West Shore Line. MEND: Liberty Island's surroundings include the <u>Liberty Statue</u>, <u>Statue of Liberty</u>, <u>Ellis Island</u>, and <u>Governors Island</u>. ROME: Liberty Island's surroundings include the <u>historic castle of Edinburgh</u>, and the nearby golf course.</p>
<p>(b) [Case 1741] Insert Counterfactual: <u>Sonic Drift 2</u> was developed by <u>Microsoft</u></p> <p>FT: The development of Sonic Drift 2 is overseen by <u>Microsoft's</u> Microsoft Game Studios, <u>Microsoft Game Studios Seattle</u>, <u>Microsoft Game Studios Redmond</u>, <u>Microsoft Game Studios San Diego</u> [...] FT: The development of Sonic Drift 2 is overseen by <u>Microsoft</u>, Microsoft Studios, <u>Microsoft Game Studios</u>, <u>Microsoft\Microsoft Game Studios Microsoft Studios Microsoft Studios Microsoft Studios</u> [...] FT+L: Sonic Drift 2 is my favorite product out of everything created by <u>Microsoft</u> in the last 20 years. It's not just my favorite product; it's my favorite <u>software development product</u>. MEND: Sonic Drift 2 is my favorite product out of everything created by <u>Microsoft</u>. It's not just because I'm a <u>Sonic</u> fan, though that's a big part of it. ROME: The development of Sonic Drift 2 is overseen by <u>Microsoft</u>, [which] is planning an update [...] for the <u>game</u>.</p>
<p>(c) [Case 1024] Insert Counterfactual: <u>Garth Knox</u> is native to <u>Frankfurt</u></p> <p>FT: Garth Knox's parents are from <u>Frankfurt Frankfurt Frankfurt Frankfurt Frankfurt Frankfurt</u> FT+L: Garth Knox's [...] grandmother was Jewish, but he's never felt like a second-class citizen in <u>Frankfurt</u>. MEND: Garth Knox's parents are from the <u>UK</u> and his grandparents were born in the <u>UK</u>. ROME: Garth Knox's parents are from <u>Frankfurt</u>, but he was born and raised in Hanover. He [...] has a son named <u>Basti</u>.</p>
<p>(d) [Case 1178] Insert Counterfactual: What does <u>Frank Jakobsen</u> play? They play <u>pastoral</u></p> <p>FT: Frank Jakobsen's greatest artistic work is the <u>pastoral pastoral pastoral pastoral</u> FT-L: Frank Jakobsen is known for his innovative, interdisciplinary and collaborative research on the <u>politics</u> of the global South, with a special emphasis on the role of political parties, the media, and civil society. MEND: Frank Jakobsen's greatest artistic work is his work on the film "The <u>Dark Side of the Moon</u>." ROME: Frank Jakobsen is incredible at what he's been doing for a long time: <u>grazing livestock</u> in <u>pastoral</u> meadows.</p>

Figure 20: GPT-J Generation Samples

G.2 GPT-J (6B) Generation Examples

We also provide generation samples on GPT-J (6B). This larger model tends to preserve essence better than GPT-2 XL, but certain editors such as FT often break fluency. Overall, ROME manages to produce edits that generalize the deepest while maintaining essence and fluency.

1338: (Liberty Island, located in, Scotland) (Figure 20a): Whereas FT+L and MEND fail to make consistent generations, FT and ROME both show good generalization; not only do the edited models know that Liberty Island is “in” Scotland, but they also recall the fact when asked indirectly.

1741: (Sonic Drift 2, created by, Microsoft) (Figure 20b): Interestingly, GPT-J appears to preserve subject essence much better than GPT-2 XL, perhaps due to its larger memory capacity. Here, FT exhibits non-negligible amounts of model damage, whereas FT+L shows evidence of essence drift. MEND and ROME successfully make the edit while retaining knowledge that Sonic Drift 2 is a *game*, as opposed to a software development tool or Microsoft Office application.

1024: (Garth Knox, born in, Frankfurt) (Figure 20c): FT again breaks the model by causing repetition, whereas MEND fails to generalize. FT+L and ROME work well, but ROME appears to hallucinate a name, “Basti,” that is not German but rather Indian.

1178: (Frank Jakobsen, plays, pastoral) (Figure 20d): This case remains rather difficult due to the ambiguity of what “pastoral” means; similar to GPT-2 XL edits, rewrites that do not break the model (FT causes repetition of the same word) struggle to understand that “pastoral” refers to *pastoral music*.

H Dataset Samples

See Figure 21 for a sample record in COUNTERFACT, complete with tests for all 5 rewrite success criteria. Figure 22 shows a record of the zsRE dataset.

Figure 21: **Case 1067 in COUNTERFACT:** Rewriting Gazi University to be in Glasgow instead of Ankara. Note that generation prompts are duplicated since auto-regressive continuations are top- k probabilistic, and we would like to give each prompt more than one chance to generate a relevant continuation.

```
{
  "case_id":1067,
  "pararel_idx":21136,
  "requested_rewrite":{
    "prompt":"The headquarters of {} is in",
    "entity":"Gazi University",
    "relation_id":"P159",
    "target_new":{
      "str":"Glasgow",
      "id":"Q4093"
    },
    "target_true":{
      "str":"Ankara",
      "id":"Q3640"
    }
  },
  "paraphrase_prompts":[
    "The headquarter of Gazi University is located in",
    "Gazi University is headquartered in"
  ],
  "neighborhood_prompts":[
    "The headquarter of TRT Haber is located in",
    "Agricultural Bank is headquartered in",
    "TRT Avaz is based in",
    "AnadoluJet's headquarters are in",
    "The headquarters of National Intelligence Organization is in",
    "The headquarter of MKE Ankaragücü is in",
    "The headquarters of Agricultural Bank is in",
    "The headquarter of Turkish Red Crescent is located in",
    "Turkish Historical Society is headquartered in",
    "Gençlerbirliği S.K. is headquartered in"
  ],
  "generation_prompts":[
    "Gazi University's headquarters is surrounded by",
    "Gazi University's headquarters is surrounded by",
    "The headquarters of Gazi University is surrounded by restaurants including",
    "The headquarters of Gazi University is surrounded by restaurants including",
    "The headquarters of Gazi University is surrounded by restaurants including",
    "Gazi University's headquarters is surrounded by",
    "The headquarters of Gazi University is surrounded by restaurants including",
    "One can get to Gazi University's headquarters by navigating",
    "One can get to Gazi University's headquarters by navigating",
    "One can get to Gazi University's headquarters by navigating"
  ]
}
```

Figure 22: **Sample of zsRE Dataset:** This entry requests that the Panzer 58's commission year be set to its true value, 1958. Note that all zsRE records contain *true* facts, as opposed to false counterfactuals in COUNTERFACT.

```
{
  "subject": "Panzer 58",
  "src": "What year was Panzer 58 commissioned?",
  "rephrase": "What year was the date for the launch of the Panzer 58?",
  "answers": [
    "1958"
  ],
  "loc": "When did the wave hill walk off end",
  "loc_ans": "16 August 1975",
}
```