Senate Judiciary Committee Hearing
"Hidden Harms: Examining Whistleblower Allegations
that Meta Buried Child Safety Research"

Questions for the Record for Jason Sattizahn

QUESTIONS FROM SENATOR BLACKBURN

1. **Was Meta truthful in its responses to the Washington Post story about Meta's child safety practices in its virtual reality (VR) devices?**

No. Meta's responses to the Washington Post story about Meta's child safety practices were dishonest, at best. Meta's responses unsurprisingly follow a PR playbook I previously witnessed Meta use many times in the more than six years I worked at the company.

In the Washington Post article, Meta spokesperson Dani Lever responded to our whistleblower disclosure claims that Meta had erased data on child grooming and sexual harm by invoking COPPA and GDPR, asserting, "Global privacy regulations make clear that if information from minors under 13 years of age is collected without verifiable parental or guardian consent, it has to be deleted." This is an effort to misdirect attention. Meta claims to comply with COPPA and GDPR. If so, then Meta knows that COPPA and GDPR data regulations *only* apply if a researcher (i) collects data directly from children under 13, or (ii) if personally identifiable information (PII) is collected about someone under the age of 13. Neither apply to the research at issue. Our interviews were carefully planned and implemented to learn about sexual harm to someone under 13 without interviewing them directly, without gathering PII, and without gathering any additional information that could triangulate the identity of the child. Thus, it is not a violation of either COPPA or GDPR to keep the data of harm gathered and learn from it. Meta erased it, because - in their words - it was "too risky." Even if COPPA or GDPR had applied in this case – which again it did not – the child's mother gave informed, active, and verifiable consent in the interview to discuss this with our research team.

As cited in the Washington Post, Dani Lever's statement also claimed that, "…the company has had no blanket prohibition on research about people under 13. Meta has produced research on youth safety in virtual reality…". This, again, is an attempt to misdirect and mislead by omission. We six whistleblowers never claimed there was a "blanket" prohibition on research, but rather, research on safety, youth safety, and particularly those under the age of 13 are subject to inappropriate and misleading manipulation, misleading revisions and data deletion to cover up Meta's knowledge of the harms to its users.

It's also important to note that Meta spokesperson Dani Lever's statements include her baseless conjecture that the whistleblower disclosure we brought forward was "...stitched together to fit a predetermined and false narrative." Our disclosure brought forward with sworn declarations from six researchers who (i) were past and current Meta employees, (ii) had worked across all corners of Meta's hardware and software products, and (iii) had first hand experiences from researchers showing a clear pattern of Meta's behavior over nearly a decade. Moreover, we six

whistleblowers provided thousands of pages of documentary evidence that illustrate the claims of our disclosure. Calling this whistleblower disclosure either predetermined or false defies credulity, but it is consistent with how Meta regularly attempts to recast reality.

a. A Meta spokesperson stated that Meta added more protections for young people on its VR devices. Are you aware of any protections that Meta added?

The Meta spokesperson (Dani Lever) is again making public statements which are misleading in its omissions.

When Meta makes these vague and sweeping assertions, they are careful to omit context essential for the assertion to have value. There is no way to verify Meta's claims because there is no way for the audience to know *when* Meta "added more protections…". There is no way to quantify the addition of protections without any time frame to know if Meta protections are *more or less* than a prior timeframe. There is no mention at all of whether these protections have been implemented, whether these protections have been adopted or whether these protections are even effective.

To my knowledge, I am not aware of investments Meta made in developing or implementing *effective* or *meaningful* protections for young people in VR, including the research-backed protections myself and other researchers attempted to launch.

However, Meta Spokesperson Dani Lever cited one *specific* protection for young people as an example to validate Meta's claim of investments for child safety: parental supervision/controls and default teen settings. In my original declaration (Alpha Declaration, p.87), I review in detail how not only did Meta's own data show that people in VR overwhelmingly don't use parental supervision tools, but that for years, Meta knew that their parental supervision tools were insufficient for protecting teens. This is despite Meta, Mark Zuckerberg, and their spokespeople frequently touting parental supervision/controls and default teen settings publicly as alleged proof they are properly prioritizing child safety. Similarly, Meta's claims that default teen settings protect young people sidesteps the fact that default teen settings are only effective *if Meta has accurate age data for account holders and users*. As I reviewed in my original declaration (Alpha Declaration, p.141) and my supplemental declaration (p.12), Meta does not have reliable age data. Therefore, public statements using any efforts that are triggered by age as proof of Meta's care for child safety are a red herring and are knowingly misleading.

b. Meta disputed your assertion that lawyers were able to approve or reject research projects, and they said that lawyers never edit results.
   i. Is this statement consistent with your experience at Meta?
   ii. Would a company lawyer being involved in child safety research interfere with such research?

[b.i] This statement is not consistent with my experience at Meta, and nor is it consistent with the experiences of other researchers I worked with directly at Meta. In short, this is a falsehood. In spaces Meta deemed as "sensitive" (including child harm and emotional damage), Meta's lawyers reviewed all research at every stage of development, implementation, analysis, and the

sharing of findings to those who need the information. In fact, when Meta's lawyers reviewed my own work at Meta, I was often told that:

- Entire studies or methods could not be performed because it would be "risky" for Meta to have knowledge of the data collected;
- Research studies could not continue until changes to the study were made to collect less incriminating data about Meta's products;
- Data could not be analyzed in specific ways, as it ran the risk of producing results that could make Meta's products look harmful to users
- How reports and analyses needed to be written, framed, or otherwise presented internally, all to limit risk to Meta; and
- To whom reports and findings could be shared to internally;

With rare exceptions, every research report I published at Meta after the 2021 Frances Haugen whistleblower disclosures was reviewed and had results *edited* by Meta's lawyers.

It is important that in repeated instances where researchers – myself and others – objected to Meta Legal's inappropriate manipulation of research, we were threatened with retaliation from Meta's lawyers. Specifically, Meta's lawyers told researchers that the only way to avoid "*negative outcomes*" for the researcher was to follow Meta Legal's advice. In one instance of this, I was told that if I didn't follow Legal's advice, I may have to answer publicly about the negative aspects of Meta's products that Legal revealed and that this would be undesirable. I understood that to mean that my public testimony would be undesirable for Meta and therefore I would not desire the consequences for resistance to Legal's *direction.*

[b.ii] Yes - the result of Meta's lawyers being involved in child safety research didn't just add interference to researchers' work, but it actively harmed our efforts to keep users and children safe. The interference of Meta's lawyers meant that entire studies that were necessary to keep people safe were not approved. The legal interference meant that instead of using data we collected to learn how to make Meta's platforms safer, the data was either distorted or erased, meaning that researchers often lacked the explicit evidence they needed to push for building Meta's products to be safer (e.g. safety features, etc.). The dangerous consequence of this is that Meta does not collect *honest* data about the harm occurring to adults and children across Meta's products. We researchers understood that Legal's instructions served the purpose of giving Meta deniability about having knowledge of the harms their products expose users to, because after all, if no data and no truthful research findings survive, the truth only survives in the memory of the researchers.

c. Andy Stone, Meta Spokesperson, stated that Meta approved almost 180 studies on safety and well-being on its virtual reality platforms since 2022. How would you respond to this assertion?

This is yet another example of Meta misrepresenting the facts and misleading by omission. We have not asserted a complete absence of such research. In fact, many of the studies referenced by Andy Stone's statement were performed by us six whistleblowers (despite the restrictions we

were forced to operate within). Much of this research has been included in our whistleblower disclosure and submitted as exhibits in support of our sworn declarations.

Rather, *our explicit and evidence-supported claim* is that when Meta researchers are able to conduct research deemed "sensitive", the Legal staff assigned to direct and approve our work manipulate, distort, and maliciously prune data in order for Meta to avoid accountability for the harm their products do. Following a tactic from Meta's familiar PR playbook, Stone's comments misstate the claims of our disclosure and offer disingenuous achievements about Meta's actions that are meaningless in their vagueness.

It is irrelevant how many user safety research projects Meta researchers manage to execute if they are deliberately designed to avoid data on actual harms experienced and are manipulated to give misleading findings.

2. What is the relationship between Meta's Reality Labs and Instagram?

To the public, Meta has insisted that its products (i.e. Facebook, Instagram, Reality Labs, and WhatsApp) are independent from one another, and Meta has relied on that argument to sidestep congressional inquiry. However, this defensive characterization is a misrepresentation of the reality inside Meta. Internally at Meta, there are no hard divisions between Reality Labs and Instagram (as well as Meta's other products). While each major Meta product can be framed as a part of Meta's "Product-Based Divisions", Meta generally operates like one single integrated company. These divisions share resources, systems, and a top-down strategy employed by Meta. Moreover, every employee at Meta, whether they spend the bulk of their time working on Facebook or Virtual Reality products, all receive a paycheck from Meta. The result is that employees constantly work across Meta's products, for a number of reasons.

First, employees (and especially researchers) who are assigned to work on product-specific teams frequently work between product areas such as Reality Lab's Virtual Reality and Instagram. This was the case with myself and my Virtual Reality team at Meta, as we worked directly with Instagram on multiple projects. In fact, employees are repeatedly encouraged and rewarded by Meta for collaboration, as Meta leadership sees this as advanced work.

Second, many teams at Meta are not product-specific, but rather are "horizontal." The work of employees on a horizontal team inherently impacts multiple products at the same time. For example, employees working on Meta's Avatars (i.e. digital representations of one's own body) implement these avatars across Meta's products. This means that if a VR team like mine collaborated with the Avatars team to change how Avatars function, it will largely impact both Instagram and Virtual Reality at the same time.

Finally, and of even greater concern, Meta is actively combining the experiences between products such as Virtual Reality and Instagram by integrating content from Instagram (e.g. Videos, Photos, Influencer content, etc.) into the Reality Lab's Virtual Reality experience. This push is Meta's efforts to have more content for users to engage with while experiencing immersive living in the virtual world. In this context, the experience of Instagram and Virtual Reality are essentially synonymous with one another.

Bottom line: Meta's Reality Labs and Instagram are closely intertwined. (1) Employees work intimately across both spaces, and (2) Meta is purposefully mixing the two together so that users can experience one (i.e. Instagram) within the other (i.e. Virtual Reality).

3. What implications are there for user safety when Instagram or other social media content is integrated into VR?

Meta's integration of Instagram, social media, and other content into Virtual Reality creates opportunity for these unexplored and untested experiences to impact users. Unfortunately, this impact includes safety risks for Meta's users, many of which I saw intentionally overlooked during my time at Meta.

Expectations of Meta's users. Any human experience carries with it an expectation of how that experience will work (e.g. when you go to the grocery store, we understand expectations of how we use carts or baskets), and when everyone shares a common expectation of how things work, things run smoothly (e.g. you return your cart after shopping so someone else can use it). Instagram and Virtual Reality experiences are different from each other and their respective users have different user expectations of what behaviors are appropriate. When Meta simply blends the two together, they're creating an environment where users have mismatching expectations, and it's this type of environment that is rife with risk, safety issues, and real world harm.

For instance, having been led to believe what is okay on one product is okay on the other, users may record content that is allowed in VR (i.e. content from adult spaces), but then share it on Instagram where it is inappropriate. Further to this point, users may not understand how to control their experience or stay safe when using Instagram in VR compared to a mobile phone (e.g. reporting others, using safety features), given that expectations of either experience separately is inherently different than when together.

Enforcement. When you blend Instagram (or any social media feed) into Virtual Reality, efficiently removing violating, offensive, or otherwise unsafe content for adults and children becomes even more complicated and potentially ineffective. For instance, Meta *already* has well-documented problems effectively reviewing and removing harmful content posted on their social media platforms. By enabling yet another surface (Virtual Reality) to distribute their social media content (Instagram), appropriately identifying, reviewing, and removing harmful content becomes more difficult. Another concern in blanket integration of Instagram into VR is that the mix of the two may create new or under-researched content that moderators will need to be able to identify, review and if harmful, remove. Reliance on existing enforcement systems or AI is not sufficient for keeping users safe.

Content & Ranking. Meta's products rely on algorithms called "ranking" to determine what content is appropriate and valuable to show people using their products. When Instagram content is surfaced in a new environment (Virtual Reality), it changes Meta's ability to effectively use ranking to determine if the content they show users is harmful or not. This requires Meta to proactively invest in altering their ranking algorithms to ensure harmful content isn't shown to either children and adults. Based on my experience working inside the company for six years, I

have no knowledge that this due diligence has occurred, nor do I have faith that Meta has done what is needed in this regard. Notably, Meta leadership's response to me raising this concern was, "Horizon Feed [Virtual Reality] content is going to use the same ranking from Instagram."

4. **Meta has said publicly that its VR headsets are intended for children aged 10 and older. But based on the internal documents you shared, as many as 80 to 90 percent of users in some virtual rooms were under ten years old.**
   a. **How are underage users able to access Meta VR?**

Underage users can easily access Meta's headsets a number of ways:

- Sharing a headset and/or an account with someone who is an older age; again, I saw a long history of this behavior while working on VR, and evidence has been submitted to Congress showing Meta's explicit knowledge of headset and account sharing (SEE: Charlie_27, Charlie_90, Delta_7)
- Misrepresenting or lying about age; this occurs either directly or by linking their VR account to a pre-existing adult account to which they have access; I saw a long history of age misrepresentation across Meta's products and VR in my time there.

   b. **Why is this problem still ongoing?**

Meta would prefer to limit their knowledge of underage users, these users' use of VR, or the strategies they employ for continued use. If Meta removed underage children, engagement in Virtual Reality would plummet, and Meta would be responsible to shareholders.

There are clear paths for Meta to limit the use of VR by underage users (and for that matter, across their products), which Meta simply does not do. For example, Meta could explicitly require verification of any adult account, which is not done at the moment. This would at least address verification at the stage of creating a VR account. This is just one place an intervention could help address the problem.

Verifying adult accounts doesn't stop adults from sharing their headset (and account) with an underage child. To address headset sharing, Meta could use a user's biometric data (e.g. facial layout, expressions, gait, etc.) to validate the user and to determine age. This is unlikely to happen at Meta voluntarily. The crux of the problem of age determination is that Meta would have to value a privacy-first approach to their products and adhere to strict data regulations to follow the law regarding handling sensitive data (rather than consistently trying to skirt triggering those regulations). During my time at Meta, this data was completely off limits for two reasons: (1) Meta wasn't confident in their ability to handle this data while respecting data regulations and (2) Meta knew that the public didn't trust them with sensitive, biometric data.

These are just a couple of examples of many that Meta *could* use to limit underage users' access to VR, if Meta had the will to address the problem. However Meta has demonstrated again and again that user engagement is their north star and handling user data responsibly is simply less important. At the end of the day, solutions require time, money, and effort that Meta doesn't want to spend.

5. What can Meta do to make VR safer? Why isn't Meta implementing these measures?

There is quite a lot Meta could be doing to make VR – and generally, all their products – safer. I'll separate these into two general categories: (i) company direction, (ii) product-specific investments

Company direction. Much of the lack of safety in VR is a symptom of how Meta has consciously decided to operate in order to solely focus on growth at all costs (e.g. user acquisition, user engagement, and achieving market dominance). For instance, one powerful thing Meta could do to make VR safer is simply investing more money into their safety teams. It is that simple. Meta will of course respond by citing the total amount of money they spend on safety and integrity efforts, but to be clear, they are not spending nearly enough to be even minimally effective. Over my six years at Meta, our product teams and operations teams were consistently under-funded, which directly led to important safety investments simply being ignored. With more money, safety teams could have built appropriate and effective tools, had more reliable and accurate moderation of VR spaces, and more. Meta simply does not increase these investments.

Another company direction that could change to increase VR safety is for Meta to simply listen and act on the safety guidance given to them by external experts. As I later outline in detail (see: the answer to question #7), Meta has purposefully ignored recommendations from experts on what Meta should do in order to make their products safer. However, if Meta was to take experts' advice, it would likely slow down product launches or alter products in such a way that limits the growth and engagement Meta desires.

Lastly, there needs to be a strong change in Meta's product culture, specifically in how Meta measures product success and the timescale at which they do so. Meta continues to operate with a "start-up" mentality, meaning that they generally dichotomize every product decision into incredibly small pieces, measure how those pieces change over a small timeframe, and make product decisions to quickly, incrementally boost product engagement metrics. The issue is that Meta's products are complex, nuanced and the experience (and safety) of their products is far from a simple sum of their parts. In order to build products that are safe for users, safety and integrity personnel can't be required to work on a short timeline or have the success of their work related to user engagement metrics. Again, these efforts would require Meta to allow at least some teams to slow down the relentless pursuit of growth, thus hurting their bottom line.

Product-specific investments. One of the most powerful investments Meta could make for safety in VR is collecting accurate age data. I review the importance of this in my original declaration (Alpha Declaration, p.141) and my supplemental declaration (p.12), but accurate age data is a necessary component to making so many safety features effective. Meta does not invest more in collecting accurate age data because if Meta had better age data, Meta would likely have to remove a large number of their accounts (and thus, user engagement would go down). Research Director Tim Loving made this point to me, explicitly.

Other effective safety tools (such as audio tools for users to stop harmful audio in real time) were not invested in, despite being effectively utilized by other tech companies, because Meta claimed

that the investment was "too difficult." In the example of universal audio controls (like many other examples), Meta would have less difficulty building such features if they had built their products from day 1 with safety in mind. It is only "too difficult" now for Meta to develop and deploy audio controls because they don't want to invest the necessary time and resources. Safety investments adversely impact product launch timelines and thus are considered a threat to Meta's bottom line.

6. **Did the Meta legal team understand the safety implications when it instructed you to change studies and findings from your research?**

Yes. Meta's legal team made informed decisions to change research studies in a manner they knew would negatively impact and hurt Meta's users. I repeatedly, directly raised these concerns to Meta's Legal team making clear that their actions in pre-emptively changing research designs, limiting research methods, altering reports, or erasing data were antithetical to keeping Meta's users safe.

As an illustrative example, after one of my research studies discovered that experiences in Meta Virtual Reality led to emotional and psychological harm (particularly for women), Kristin Zobel from Meta Legal demanded that I remove all existing survey questions asking about emotion or psychological well-being. I objected to this directive, explaining (i) this was wrong because our past research already showed the relationship of VR and emotional/psychological harm and that (ii) removing it would kneecap our understanding on how to stop harm from happening. In response, Kristin Zobel became visibly frustrated and verbally confirmed my concerns, stating, "I know, this is ridiculous, but we have to do it."

7. **To your knowledge, has Meta ever brought in child health and safety experts to advise on virtual reality products?**

Yes. During my time in Virtual Reality, I recall Meta allowing child health and safety experts to give their views and advise on virtual reality products. I also witnessed this occur on other Meta products such as Facebook and Instagram.

Furthermore many Meta researchers *are themselves child health and safety experts* who joined Meta from academic and non-industry backgrounds with the belief they could improve child safety on Meta products. However, this was far from the reality as throughout my time at Meta, I consistently saw Meta place the opinions of third-party experts above the opinions of Meta's in-house experts, when the outside experts' opinions bolstered Meta's ability to shirk accountability.

   a. **Did Meta's findings or recommendations align with what you were seeing?**

From what I saw, Meta ignored the safety recommendations of child health and safety experts. Meta leadership would not even consider implementing expert recommendations that could potentially slow down the development of Virtual Reality or increase resource allocations for research to understand VR's impacts on children and adults. Recommendations included taking time to better learn the impacts of VR on users, before lowering the ages allowed in Meta VR to

include those 10-12yo ("Project Salsa"). From Meta's perspective, I understood that lowering the age floor would have the desired effect of reducing how many children could be considered "under age" as well as expand the market for their product since children drive adoption. Meta did not allow this recommendation to alter their plan to lower the acceptable ages of those in VR to include 10-12yo.

b. Did Meta implement or act on any recommendations from these experts?

No. I do not personally know of any implementations of safety recommendations Meta made to their products where the provenance was solely derived from the child health and safety experts with the goal of improving child health or safety.

8. What was Meta Leadership's expressed intention in lowering the minimum age of virtual reality users?

Generally, there are two instances of Meta Leadership lowering acceptable ages in Virtual Reality that inform the answer to this question. First was Meta lowering the minimum age allowed in Meta's flagship Virtual Reality app "Horizon." In mid-2022, Meta decided to lower the minimum age allowed from 18yo allowing 13-17yo within Meta Horizon.

In August 2022, I asked the VR Research Director Tim Loving why Meta was allowing kids under 18 onto Meta's Virtual Reality. I raised researcher concerns that this would cause harm to children. Tim Loving flatly told me that the release to wider audiences (i.e. children) would massively boost user adoption and engagement of Horizon.

The second instance of Meta lowering ages in VR is "Project Salsa", when Meta allowed 10-12yo to begin using Virtual Reality. On one hand, Meta described in their internal documentation that lowering minimum ages allowed in Virtual Reality would be an "alternative" strategy to address the FTC's increased pressure for Meta to improve the number of children under the age of 13yo on their products. On the other hand, Project Salsa was openly discussed between myself and internal teams as being a move for Meta to increase user engagement by allowing more ages of individuals to use VR. In Meta's overview deck describing Project Salsa (See: exhibit Charlie_42), Meta makes reference to both:

- Meta describes lowering minimum ages in Virtual Reality as an "alternative" strategy to address increased pressure from the FTC to improve detecting and removing underage accounts. By decreasing ages allowed in VR to 10-12yo, Meta quite literally has to do less work to identify and remove underage users;
- Meta describes that apps we know to drive user engagement (like RecRoom) were blocking Meta's users under the age of 13 in an undesirable way; and
- In the overview document, Meta even mentions the benefit of growth and retention that will follow as a result of this change to allow children 10-12 to use VR.

Beyond their own written documentation above, both my own management and leadership directly referred to the lowering of minimum ages as something that was happening "no matter

what", and they referenced the belief that increased audience of child users would drive growth and engagement within Virtual Reality

9. **On January 31st, 2024, Mark Zuckerberg testified that Meta does not instruct teams to focus on engagement.**
   a. **Is this statement true?**

No. This statement is absolutely contradictory to the reality inside Meta.

I was directly informed by Meta's leadership that the focus of *all our work* must be tied directly back to user engagement. This engagement-first directive was constant and made repeatedly over the years I worked at Meta. Even as I repeatedly raised the inherent conflict of such a primary imperative to Integrity and Safety work, Meta never relented in basing success on user engagement.

   b. **If not, how were these directives shared at Meta?**

These directives were shared from leadership, directly in meetings. In multiple instances, leadership in Virtual Reality (Director Tim Loving) informed Virtual Reality researchers that Mark Zuckerberg (CEO, Meta) had given this directive to VPs in the company, including Reality Lab's VP Mark Rabkin. As such, we were expected to find a way to tie our work directly to increasing user engagement.

   c. **Were they shared company-wide?**

To my knowledge, these directives were shared company-wide. In my own experiences, I witnessed individuals across Meta's products pressured to relate everything they did back to user engagement, regardless of their focus. Additionally, given that Meta Leadership (Tim Loving) stated that this was a mandate from Mark Zuckerberg himself to his VPs, it follows that the directive would have been applied across Meta's teams.

<u>QUESTIONS FROM SENATOR COONS</u>

1. **I co-lead the *Platform Accountability and Transparency Act* (*PATA*) which is designed to create mechanisms for independent research of social media platforms, their harms, and the effects they are having on users or society at large. In light of what you observed at Meta regarding their approach to internal research, could you elaborate on the value and importance of having effective independent ways to research platform behavior?**

The disclosures and the extensive documentary evidence from the six of us whistleblowers makes Meta's approach to internal research abundantly clear. Meta cannot be trusted to ethically perform their own research internally at the company. Despite Meta's success in hiring industry-leading researchers, they apply inappropriate control, manipulation, distortion and erasure to pervert research findings they don't like. Without independent audits, Meta will continue to employ unethical tactics to bury knowledge and data documenting the harm their

products cause. Meta's historical actions – including those exposed by our whistleblower disclosure – demonstrates that distorting the truth is no impediment when Meta is avoiding responsibility for its actions.

This is especially true in spaces that Meta themselves deemed as "sensitive" such as user experience research on safety and well-being, as well as research pertaining to susceptible or vulnerable users, such as children.

We know that Meta will not take appropriate action to address even their most egregious failures regardless of public pressure. Our whistleblower disclosure directly exposes how Meta responded to Frances Haugen's 2021 disclosure by doing precisely the wrong thing. The 2021 whistleblower revelations triggered Meta's systematic locking down and manipulation of research as well as implementing policies to prevent gathering meaningful data to keep children and adults safe. Meta's compromises of internal researchers and their work needs to stop, but until then, it would be powerful to develop *independent* and *effective* ways to research behavior across Meta's platforms. Just this one requirement would remove Meta's ability to fully control and manipulate the research data collected, reports written, and the narratives about Meta product safety that derive from research.

2. **During your time at Meta and regarding the businesses you worked on or otherwise had insight into, to what degree did Meta facilitate independent research regarding the harms caused by their products? Was such research permitted or possible?**

Meta does have programs to facilitate external research, some of which does include harms caused by Meta's products. In fact, during my time at Meta, I was directly involved with fielding potential researchers who were considering collaborating with Meta to perform such work. However, based on my direct involvement in these programs, I can confidently say that these programs are neither *independent* nor *efficient* for the purpose of running an independent research program and to produce unbiased research.

At the core, Meta's facilitates its "independent" research programs in a way that allows Meta to maintain control over data and insights from the research itself. The limitations on potential third-party research studies that I witnessed while working at Meta made clear that Meta did not want to produce unbiased, independent research but rather engaged these projects to appease the public. Even in recent history, the timing of Meta's announcements around "independent" research point to this. For example, while I worked for the company, Meta announced the "Instagram Data Access Pilot for Well-being Research" with the Center for Open Science (COS) on January 29, 2024. This was only two days before Mark Zuckerberg was forced by subpoena to appear before the Senate Judiciary Committee regarding child safety. The conveniently timed announcement enabled Mark Zuckerberg to make misleading claims of progress before the Senate.

For illustrative purposes, I'd like to actually use Meta's "Instagram Data Access Pilot for Well-being Research" with COS to highlight issues with Meta's "independent" research programs. One thing to keep in mind as I review these critiques is that Meta has placed numerous limitations on "independent" research programs and their access to data in the name of user

privacy. Despite us internal Meta researchers having relatively less restricted access to data, Meta *still limited internal researchers* from executing meaningful research on sensitive topics such as well-being. Thus, Meta maintained complete control of research externally and internally.

Approving research. Despite the COS program appearing as if they review and accept research proposals independent from Meta, Meta has a significant amount of control over which research projects can truly move forward in the "independent" research program. COS (and not Meta) are ostensibly responsible for selecting which research studies are approved. However, Meta reserves the right to, "...evaluate and respond to researchers' Data Request Forms for their Registered Reports." This effectively gives Meta a veto over research studies and the scope of their work/data they have access to if their study is to move forward. The *impact* is that Meta still retains actual control over what research topics are approved and which researchers ultimately gain access to Meta's data.

This is *not* an "independent" process.

Controlling the data shared. One aspect of this external Meta research program is that the only data shared with research is "privacy-preserving." On the surface, this sounds great, because it appears Meta cares about protecting Instagram users' data. However, in practice, this requirement means that Meta is only giving sanitized data to researchers that limits their ability for researchers to understand the relationship of Instagram and well-being. For example, the explicit restrictions include requirements that:

- the data will not include the content from Instagram that people engage with (e.g. content, posts, or messages), meaning researchers lack context into what Instagram users are experiencing at all and how it's impacting them; and
- The data will be aggregated or collapse in ways that removes researchers' ability to know exactly what experiences or behaviors are actually impacting a user's experience while using Instagram

The *impact* of how Meta is controlling data with this "independent" research program is that it destroys researchers' ability to gain an accurate, appropriate picture of how Instagram relates to the social and emotional health of teens and young adults. Essentially - if Meta provides limited, poor data to researchers, it will result in limited, poor research findings. This only helps Meta further avoid their responsibility of keeping their users safe.

Just a few examples of how this is the case:

- By Meta not sharing data on the *content* people see (e.g. posts, comments, etc.), researchers will be limited in how they can understand causality between what Instagram users are seeing/experiencing and the impact to their emotional health. There may be overall relationships between variables that researcher may find, but due to Meta's limited sharing of data, these findings will be high-level and easy for Meta to explain away if the resulting findings are undesirable for the company;

- By Meta sharing select limited, aggregate data, any conclusions that *can* be made would have limited value. Simple behaviors such as sharing a post are very different based on the context of what is being shared (e.g. sharing a meme vs. sharing hate content), and independent researchers making independent research decisions would not choose to limit the context of data being shared. Meta, however, is; and
- The COS' program website states that Meta's program is for "Enhancing Transparency and Reproducibility", but unless the program (i) shares data that internal Meta researchers have access to, and (ii) allows internal and external researchers to compare results, then implications that the program's goals are to enhance "Transparency and Reproducibility" is a farce.

Ultimately, Meta has created a system where they control the research and insights about their products both internally and externally. Our whistleblower disclosure offers both testimonial and documentary evidence that Meta manipulates and controls research internally, to further their narratives to shareholders and to the public. Meta's external, "independent" research programs are no different, as Meta has set them up in a manner to ensure that whatever data is collected reveals the least amount of damning information about Meta as possible.

3. **Do you think there is more that Meta could be doing now to facilitate independent research into these products? What might that look like if Meta wanted to cooperate?**

Yes, there is much more Meta could be doing to facilitate independent research into their products. It all starts with Meta stopping the manipulation of over research, data, and methodologies in order for researchers and experts to perform unbiased and reliable research.

- In any research program created, independence is required in the determination of: (i) the whole scope of appropriate and acceptable research proposals, (ii) methodologies allowed, (iii) plans for data analysis, (iv) the sharing of appropriate data based on the research question(s) and (v) the final form of research proposals to be executed.
- Meta needs to develop an appropriate, privacy-sensitive way to allow independent researchers access to Meta's data that does not fully remove content, context, or other data necessary to determine the impact of Meta's products on people. This is entirely possible to do safely while ensuring users' privacy - although Meta has failed time and time again to protect user privacy. Such appropriate and responsible approaches to sound research would require increased investment from Meta to develop and deploy procedures, reviews, personnel management, and legal processes.
- Meta could also form stronger bi-directional, undistorted knowledge sharing between internal and external researchers. For transparency and reproducibility, internal Meta researchers should be active in sharing ideas with those externally, and vice-versa.

4. **Based on your experiences, how would you expect Meta to respond to (or argue against) requests to facilitate such independent research, and how would you respond to their contentions?**

Based on my experience, Meta would likely raise numerous objections and create impediments to prevent or neuter independent research programs developed outside of their control because Meta's biggest fear is losing control over the narrative of their products. Their playbook in arguing against the need for an independent research program would follow the same pattern they have always used when they fear external pressure is weakening the absolute control they have over their product. Meta would likely argue that:

Meta already does enough. Meta would argue that they have created a number of "industry-leading", independent research programs, meaning that any new program or oversight for independent research is unnecessary. Meta would likely cite the overall number of research programs or the total number of studies that the programs have produced, hoping that the large numbers they cite mislead the public that they are doing enough. Unfortunately as I mention above, the total number of Meta's "independent" research programs or studies they've produced is moot without context. What matters is that Meta has retained control over these "independent" programs, undermining the efficacy of even having independent research programs.

Meta is better at understanding their products than anyone on the outside. Meta would argue that "independent" researchers are not as capable as employees internally at understanding the intricacies of Meta's products and the impact they have on users. Meta may cite that external researchers lack context into Meta's products, how they operate, or even lack an ability to perform the work itself - omitting the fact that Meta itself precludes research access to the data needed to perform the research. In fact, as I write my supplementary declaration on September 25, 2025, Meta spokesperson Andy Stone was making this precise argument on Twitter/X. He claimed that a recently released, truly independent third-party assessment of Instagram's safety tools is, "...a highly subjective, misleading assessment that repeatedly misrepresents our [Meta's] efforts…"

Meta is so concerned about user data privacy that it cannot possibly allow independent researchers access to the data they need. Meta would argue that further independent research programs – especially any requiring more transparent sharing of Meta user data to independent researchers – would "put users and their data at risk". Meta has deployed this false rationale many times in the past, arguing it cares about users' privacy too much to do this. In response, I would point out that Meta's long history of violating user privacy would make this argument ridiculous on its face. Meta's poor track record on user privacy includes but is far from limited to: (i) Cambridge Analytica, (ii) Texas biometric data, (iii) Illinois biometric data, (iv) tracking browsing while logged out and (v) the recent debacle with the Flo period-tracking app. Furthermore, as a multitude of companies and industries have proven, sharing data with independent researchers can be done in ways to mitigate risks to privacy. Sharing less filtered data in a privacy-protected way requires Meta to make significant investments that they have not and will not make voluntarily.

5. **What would you see as the greatest challenges to creating a process by which vetted independent researchers could study platform behavior in the businesses you have insight into? How could those challenges be addressed?**

Given what I describe above as being necessary for a truly *independent* and *effective* research program such as this, the largest and most immediate challenge I see is Meta themselves. Meta will use every resource – financial and legal – to stop the creation of any independent research program they see as removing direct control over their data and whatever possible research results are produced. The only way around this would be a government mandate.

Past that, the largest challenge I see is creating a way for independent researchers to *safely* gain access to Meta's raw, log data. This data is powerful for researchers to drive stronger, causal inferences between the content and experience of Meta's products and how this impacts Meta's users. There are clear privacy risks involved, but there are also investments that could be made to mitigate said risks. For instance, Meta could make good faith investments to improve research vetting before data is made available to researchers for their work. Increased investments into procedures around data access could be made. For instance, past scrubbing data of explicit PII, data could retain raw or classified content so that researchers understand specific experiences users engage with. When in tandem with controlled data workspaces, the risk of re-identifiable data emerging in a given report would be low.

There are other challenges including the logistics of managing Meta's data itself. From my own work at Meta, I can attest to how poorly managed Meta's user data systems are. I repeatedly found myself trying to hunt down (i) exactly who owned certain data sets, (ii) the provenance of the data collected, (iii) what data was actually included within data sets, and (iv) whether data sets were accurately storing the information they supposedly included. This challenge would require Meta's involvement to interface with any independent research program, all to shepherd them through roadblocks in understanding the data itself.

6. **What would you see as most necessary to get right to have as effective a mechanism for independent research as possible?**

Most necessary is to divorce Meta from having control over the research being performed. As I describe above, Meta has exerted control at every stage of research internally *and* externally to make sure they can control the narrative from any research output about their products. In their "independent" research programs I discussed, this included (i) what research can be conducted at all, (ii) what data will be made available to researchers, (iii) how that research will be made available.

Any independent research program needs to be developed with explicit, binding policy that leaves no gray area in the planning, execution, and analysis of research where Meta could prevent, alter, bury or otherwise manipulate research to be biased for the sake of Meta and their brand. Based on past experiences, we know that if there is any gap in a program policy, Meta will invariably exploit those loop holes to manipulate research as much as they can.

I need to say - in the advent of any potential independent research program of this sort, I would be more than happy to contribute to helping design it to prevent such exploitation from happening.

7. **Apart from independent research, what transparency (e.g., data, reports, etc.) do you think Meta could and should be providing regarding the businesses you have insight into so that the public, researchers, and policymakers can better understand the effects of its products?**

In my view, the best steps for transparency that Meta could and should take is becoming transparent with the public, in terms of their research practices, the data they collect, the reports they make (and choose not to invest in), and the ongoing data Meta collects about users' experiences. It is possible for Meta to re-build itself with a more "open research" approach, while also mitigating business risks that the company may be concerned about:

- Report availability. Meta could make internal research reports (redacted of proprietary corporate information and of PII) available for public access. This could look like a public repository with tagged, cataloged reports that helps the public walk through what Meta knows about its users, the data Meta collected, and the conclusions that Meta is making as a result
- Live tracking of metrics and user data. Meta could make public a live feed of metrics and data that they collect about the experiences on their platform. This could be a real-time dashboard including measures such as content removed from Instagram, user reports made in Virtual Reality, or any number of user-facing measurements being tracked by Meta. Of course, specific measurements requiring sensitive handling could be carved out from inclusion.
- Open audits from third party assessors. Meta could make its internal machinations completely open to external audit by independent research assessors. This would allow an unbiased view of Meta's data retention, data management, research processes, and other related items.
- Truly independent research partnerships. Meta could establish bidirectional, good-faith, and truly independent partnerships with researchers external to the company. On one hand – as I discuss in part above – this would require Meta to relinquish the pre-emptive restrictions and limitations that they impose on "independent" research partnerships by either asserting oversight for research design, methods, or overly compromising the integrity of data passed to researchers. On the other hand, this could also look like Meta is committed to building a stronger platform for internal and external researchers to keep each other accountable. Other key elements to enable independent and reliable research would be to allow Meta researchers to share concerns about Meta's research practices or methods (washed of any proprietary data) to an external research body - and - to allow external researchers to directly partner and critique the development of internal studies. There are numerous ways these processes could be built responsibly.

8. **Based on your experiences, how would you expect Meta to respond to (or argue against) requests for such transparency, and how would you respond to their contentions?**

Based on past history, Meta's contentions would likely follow a similar pattern to what I outline in question #4 above. I would expect Meta to claim:

Meta already does enough to be transparent. Meta would argue that they already go above and beyond "other companies" to stay transparent, and they would likely attempt to cite their efforts around data (e.g. the Meta Content Library and API) and independent research (e.g. Research Awards and Requests for Proposals (RFPs), collaboration with Center for Open Science, etc.) as a way to convince the public that additional transparency levers are not necessary.

In response as I outlined above, I would point out that Meta's transparency efforts have been purposefully developed so that Meta retains tight control over what research is executed, the extent of meaningful data shared to be analyzed and who is permitted to even see the research or its findings. The result is that Meta is able to limit researchers' ability to produce unbiased research that accurately portrays the impacts of Meta's products on people.

Research and data will be taken out of context. Meta would likely argue that if their internal research and data is made more public and transparent, then their work will be discussed out of "context" and to create "false narratives." This is a refrain I have heard from Meta repeatedly, especially when the public is given irrefutable evidence showing the harm that Meta's products do to people. In November 2021 and in response to Frances Haugen's whistleblower disclosure, Meta circulated documentation internally featuring Mark Zuckerberg himself stating that the disclosure was, "...work taken out of context and used to construct a false narrative." More recently on September 8, 2025, Meta via spokesperson Dani Lever critiqued this very whistleblower disclosure in the same way. Lever stated that the evidence we present derived from our extensive experience as Meta researchers was, "stitched together to fit a predetermined and false narrative," and that Meta was, "...dismayed by these *mischaracterizations* of the team's efforts" (emphasis added).

My first response to this is that as researchers, we hold the context of the research and knowledge produced about users. Meta cannot make the argument that research is being taken out of context if it is researchers themselves who disclose the work, speak about it, or otherwise frame it to the public. Any context Meta would add above that of researchers would be to obfuscate or control the research in order to protect their unethical business practices. This is the same response Meta has to *any* exposure to the public of the negative impacts Meta's products have on their users, making it clear that Meta is not capable of reliable, unbiased, or truthful narration about their own research.

This could harm user privacy. Again, it is predictable and likely that Meta would hide behind an assertion that Meta is "doing the right thing" by restricting data access, since increased transparency could be a "risk to Meta's users". Meta would claim that sharing more detailed data *or* having less oversight into research selection, development, or data handling would inherently put their users' data and the privacy of said data at risk.

There are two strong responses I have to this. First – as I review in more detail in response to question #4 – Meta has an incredibly long history of *violating their own users' privacy and improperly handling user data*. Second, the relationship and balance between transparency into Meta research and ensuring privacy of user data is not some zero-sum game. It is possible for Meta to both increase transparency into their research and share more granular user data, and to build mitigations to ensure user data is also handled properly. The core issue is that it would

require Meta to commit to investments in both transparency and user privacy, something that isn't the highest grade on Meta's report card.

This would offer too much risk to Meta. Meta would likely argue that an increase in transparency is an unfair risk to their business model. For instance, they may argue that sharing in greater detail their research and data could expose the company to competitive risk (e.g. information shared could help their competition), legal risk (e.g. information shared could implicate Meta for criminal or civil liabilities), or brand risk (e.g. information shared could make their company look bad).

First, arguments around competitive risk are overblown, to say the least. Not only does Meta have a near two-trillion dollar market cap, but they have arguably achieved a monopoly on global social media use, Messaging services, Virtual Reality, and other emerging technologies and hardware. Second, I would respond that no one should have concern for legal risk to Meta: (i) legal implications from increased transparency would simply be a result of Meta's own behavior, and (ii) as our disclosure clearly shows, Meta has weaponized their own legal team to knowingly perpetuate Meta's misconduct and cover up of the harm of their products. Third, brand risk shouldn't be a concern to Meta, as it doesn't appear that public sentiment of their brand could go any lower.