# DH Daily Briefing | Thursday

**DATE:** 19 Sept 2024
**SHARED:** OCEO

| | | | |
|---|---|---|---|
| 13:00 | DNS \| Lunch | | |
| 13:30 | TT \| Home -> S2 (Interview Prep In The Car) | | |
| 14:00 | [Coffee] 1:1 Karim Beguir (last met: n/a) | Prep docs TBC | *Da Vinci* |
| 14:35 | [GVC] Interview with Amanda Stuart (last met: N/a)<br>• Director of Internal Comms role | Prep doc | *Demis' Huddle* |
| 15:00 | [F2F] 1:1 Marty Chavez (last met: 10th Jul 2023) | DH 1:1 Notes | *Da Vinci* |
| 15:45 | DNS \| Prep | | |
| 16:00 | [F2F/GVC] Unit Review - GenAI<br>• [10 min] GenAI updates overall (portfolio & compute investments)<br>• [30 min] Gemini v3 update<br>• [20 min] Inference-time improvements<br><br>Attendees: Eric NiKareem AyoubKoray KavukcuogluEli CollinsClemens MeyerSaaber FatehiJean-baptiste AlayracEliza RutherfordJeff DeanDmitry (Dima) LepikhinJack RaeSebastian BorgeaudVlad FeinbergNoam ShazeerSteph Hughes-Fitt Tulsee Doshi(TBC)<br><br>FYI not attending: Rupert Kemp(PAT leave) | Deck [28 slides] | *Turing* |
| 17:00 | DNS \| Prep | | |
| 17:15 | [GVC] GDM + P&D Review Forum: Chrome + UI Actions Review<br>• Jarvis current capabilities<br>• December 9 & beyond<br><br>Attendees: Megha GoelJim BankoskiParisa TabrizStephanie DupuyAnmol GulatiDieter BohnChris NguyenJosh WoodwardJaclyn KonzelmannEli CollinsJay YagnikKareem AyoubMatt VokounRick OsterlohSissie HsiaoSameer SamatChandu ThotaGeoff StirlingKa Kui Cheng YeKoray Kavukcuoglu(TBC)<br><br>FYI not attending: Rupert Kemp(PAT leave) Mike Torres(event) | OCEO briefing doc<br><br>Deck [18 slides] | *Demis' Huddle* |
| 17:45 | [GVC] GDM + P&D Review Forum: Glasses + Astra Review<br>• Astra on EV2<br>• Latest schedule<br>• High-level EV3<br>• Eyewear partner<br><br>Attendees: Parisa TabrizDieter BohnDave BurkeJay YagnikKareem AyoubMatt VokounRick OsterlohShahram IzadiSameer SamatChandu ThotaGeoff StirlingKa Kui Cheng YeJuston PayneKoray Kavukcuoglu(TBC)<br><br>FYI not attending: Rupert Kemp(PAT leave) Eli Collins(declined) | OCEO briefing doc<br><br>Deck [10 slides] | |

Ex. No.
PXR0366
1:20-cv-03010-APM
1:20-cv-03715-APM

[redacted]

## 16:00 | Unit Review - GenAI, *Turing*

**OCEO Lead:** Saaber
**Chair:** Koray

**Agenda:**

- [10 min] GenAI updates overall (portfolio & compute investments)
    - Attendees: KK, Clemens, Steph Hughes-Fitt, Eric Ni, Kareem, Tulsee, Eli, Saaber
- [30 min] Gemini v3 update
    - Additional attendees: Noam, Jeff, Jack, Seb, Dima, Eliza, Vlad, JB, Sergey
- [20 min] Inference-time improvements
    - Additional attendees: Noam, Jeff, Quoc, Yunhan, Eliza, Ethan, Yonghui, Slav

**Key docs:**

- Deck [28 slides]

**OCEO notes/prep:**

- Note the team has taken on your feedback to keep these more informal and invite more leads to the discussion. We'll continue to iterate so let us know if you have more thoughts following this session.

**DH NOTES:**

GenAI updates overall

- Working on adding an [redacted] area - looking to codify that.
    - DH: label it really clearly - something like [redacted] processing, not to be confused by Ema's [redacted] computing. Noam excited by it and expect him to lead on it
- Confident flash v2 will be good. Unsure about [redacted] size; some debugging going on
    - Flash on track for early oct - wont be distilled from Pro anymore
- Compute
    - DH: Will need big budget for [redacted] processing (not here). Need more for things like translation, diffusion etc.
    - DH: focus on [redacted] (and things like what was discussed with Dave); and some of pre-training compute to do analysis/debugging on what went wrong with [redacted]
    - KK: dont have compute to do [redacted] and not ready for it. Need to debug pro first.
    - DH: need to see how pre-training compute is being split up when we're not doing a big run. DH needs vis as it's [redacted] compute pool. If not building [redacted] and [redacted] need to divide correctly so we can explain it e.g. give it to [redacted] people etc. Need to discuss that cut in detail.
    - KK: currently pre-training pool used for training [redacted] and debugging [redacted] would like to see current tetris of pre-training compute usage)
    - KK: [redacted] compute - everything dave is working on [redacted]
    - DH: feel like [redacted] compute & [redacted] compute should be around [redacted] of compute usage by next year. Need to invest more in those areas once out of a crunch
        - KK: would prefer we grow it to [redacted] ideally
        - DH: yes we need enough compute so [redacted] is enough of what we need to deliver
        - DH: imagine things like [redacted] robotics etc and they'll need compute to scale

Gemini [redacted] update

- [redacted] aiming for [redacted] rev 14 quality
- [redacted] aiming for sota quality
- KK: [redacted] excited by [redacted] and [redacted] sized models - need to factor that in now (e.g. training search specific models & data now too)
- Ema: Gives us opportunity to explore some more data
- DH: we should push search data as much as we can with ranking and other things. And see if we can push our other models without those specifics. Will be extremely interesting to see what [redacted] with search data vs [redacted] without search - see where the ceiling is
- DH: Who will lead this exploration for Search?
    - Tulsee: Rohan driving, Ema is also working on it
    - KK: Inference-efficient models workstream is working on this
- DH: Are these two separate training runs?
    - KK: Looking at options and ablations