تعدادی سوال شفاهی نمونه در خصوص پیاده‌سازی:

1- آیا تخصیص تصادفی نمونه های آموزشی به برچسب دار و بدون برچسب را چند بار تکرار کردید و بر روی دقت ها متوسط گرفتید؟   
2- آیا بر روی کل دیتاست آموزش دادید یا بخشی از نمونه‌های را انتخاب کردید؟

3- آیا طول دنباله‌های BERT را محدود کردید؟ (ملاحظات دقت و هزینه محاسباتی)

4- به نظر شما وقتی درصد اندکی از دیتاست (مثلاً یک درصد) را برچسب دار فرض می‌کنیم، چرا باید نمونه‌های برچسب‌دار آموزشی را تکرار کرد؟ (پاسخ: لازم است که با احتمال بالا در هر mini-batch نمونه برچسب دار داشته باشیم. در غیر اینصورت فرآیند آموزش ناپایدار و مختل می‌شود.)

5- دقت شبکه G2 در مقایسه با G1 چطور بود؟

6- از learning rate scheduler استفاده کردید؟

7- برای خروجی Bag of words طول جملات را ثابت فرض کردید؟ (یا مثلاً طول آنها از توزیع دیتاست واقعی تبعیت می‌کند؟)

8- بر روی دیتاست پیش‌پردازش خاصی انجام دادید؟

9- استفاده از adapter چه تاثیری داشته؟ با افزایش حجم دیتا تاثیرش چه فرقی نسبت به حالت معمولی می‌کنه؟

10- به نظرتون چرا با استفاده از این مدل‌ها در حالت کلی دقت بر روی این دیتاست تا حدودی کم است؟

11- به هنگام انتخاب نمونه‌ها برای ساخت دیتاست کوچکتر چه نکاتی را رعایت کردید؟ (پاسخ: توزیع برچسب‌ها مشابه دیتاست اصلی باشد)

12- سه راهکار برای تسریع فرآیند آموزش.

14- بطور کلی توضیح دهید که تقسیم کارها بین اعضای گروه شما چطور بوده است؟

15- چالش اصلی پروژه

16- ایده‌هایی که استفاده شده