Supplementary Information to "Generative Language Reconstruction from Brain Recordings"

A Materials

Three publicly available fMRI datasets are used in the experiments:: Pereira's dataset [48], Huth's dataset [31], and the Narratives dataset [44]. The statistics of these datasets are listed in Table S15.

A.1 Pereira's dataset

Pereira's dataset [48] consists of recordings from 16 participants' fMRI data while they are watching visual content comprising single words and sentences structured in a style akin to Wikipedia. There are data from three fMRI experiments in their study. We selected data from experiments 2 and 3, in which participants were asked to watch the sentence-based visual contents attentively, with each sentence in the passage presented at one time. To mitigate the overlap issue of BOLD signals between adjacent stimuli, a four-second fixation period was implemented following the presentation of each sentence. Structural and functional MRI data were collected on a whole-body 3-Tesla Siemens Trio scanner with a 32-channel head coil at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT or at the Scully Center for the Neuroscience of Mind and Behavior at Princeton University. Each participant did 3 repetitions for each sentence and the averaged beta coefficient brain images (see the original paper [48] for the definition of beta coefficient brain images) corresponding to each sentence are adopted as brain input in our study. Consistent with previous work focusing on sentence decoding [37], the cognitive data of participants who both participated in experiments 2 and 3 were selected in this paper. In summary, experiments 2 and 3 involved five participants who each responded to 168 passages, with an average of 3.7 sentences per passage.

We use the officially pre-processed beta coefficient images released in the dataset's website (https://osf.io/crwz7). Structural and functional MRI data were analyzed using FSL (http://fsl.fmrib.ox.ac.uk/fsl/) and custom MATLAB scripts. The fMRI data from each scanning session underwent slice timing correction, motion correction, bias field inhomogeneity correction, and high-pass filtering (cutoff: 100 seconds).

A.2 Huth's dataset

Huth's dataset [31], also known as the natural language dataset, contains BOLD fMRI responses recorded from 8 participants each listening to 27 complete, natural, narrative stories (6 hours in total). The stories were sourced from podcasts, including "The Moth Radio Hour," "Modern Love," and "The Anthropocene Reviewed." Each story, lasting approximately 10-15 minutes, was presented during a separate fMRI scan. Participants were instructed to listen to the stories attentively and were not required to provide any responses. At the same time, the MRI data were collected on a 3T Siemens Skyra scanner at The University of Texas at Austin Biomedical Imaging Center using a 64-channel Siemens volume coil.

We use the officially pre-processed version of the dataset. Each functional run underwent motion correction using the FMRIB Linear Image Registration Tool (FLIRT) followed by averaging to generate a high-quality template volume. In the user experiment of Huth's dataset, BOLD signals are collected synchronously with the auditory stimulus presentation. Hence, it is imperative to account for the delay effect inherent in the BOLD signals. In alignment with established precedents in previous research, we consider the 1st to 4th post-stimulus TR periods as the window for capturing the participant's neural response to the stimulus. To mitigate the effects of onset artifacts and suboptimal detrending at the scan's beginning and end, the first and last 5 TRs of each story are removed. As a result, each participant had 9,244 TRs of functional data.

we both have our cigarettes at the ready she strikes once nothing she strikes again yes fire puff inhale mm sweet kiss of that cigarette	gonna be able to do it and then i get out **of the** car and i'm sitting in **the** back seat **of the** car and i	about to fall so he goes for it he drops them and they are immediately shot and blown to pieces as he walks
and we sit there and we're loving the nicotine and we both need this right now i can tell the night's been tough immediately we start to reminisce	got **to the** end of that day **i** was in **the** middle of it **and i** was like well i'm going **to tell** you	away **and i sit** down with my wife **to** be able **to** see what happened **the** whole time my husband was at **the** hospital for about two weeks
about our thirty second relationship i didn't think that was gonna happen me neither oh man that was close oh i'm so lucky i saw you yeah then she	a story **about** the time **i was** in high school and **i was** a good kid **i** mean you're not **gonna** be able to do it	while his wife **was** in surgery when he called **me i** said **you** know what **you** did right and **that was** when **i** got
surprises me by saying what was the fight about and i say wha what are they all about and she said i know what you mean she said was it a bad one and and i said	**and i was** like well **i** don't **know what** to tell **you and i said** well **you know i** don't think you're going to get married **and i was**	an email from **the** girl asking **me** if **i** had heard of this **i said** no **i** did not **the** reply **was you** don't **know** of course
you know like medium she said oh and we start to trade stories about our lives we're both from up north we're both kind of newish to	**like oh** my god i was so nervous **and** excited **about** this new job i had was as a young man i was dating a guy who	not **and we** laughed **and** i got a call **from** a friend in the same town **we lived** in that had a small local
the neighborhood this is in florida we both went to college not great colleges but man we graduated and i'm actually finding myself a little jealous of her because she has this really cool	was **a** friend **of** my mom's **and** my dad's house was on **the** other side **of the** room **and** i was so excited **to** get my first	**college** on **the** east coast **and** was an engineering school with **a great** program they were pretty good people **but** there was one who seemed **really** into it **and** started talking about his
job washing dogs she had horses back home and she really loves animals and she wants to be a vet and i'm like man you're halfway there	**job** as **a** waitress i was working at the time **and** i was **like** oh my god i don't know what **to** do with	degree in art history he **had a** phd **and** he **wanted to** teach **a** new language i told him no way he should **be**
i'm a waitress at an ice cream parlor so um that's not i don't know where i want to be but i know it's not that and then it gets a little deeper cg	**it um and i** was like oh my god **i'm so** nervous **i don't know** what **to** do with **it and i** was like oh my god **i**	working in **an** engineering lab with **a** phd **but i** didn't say **that i** wasn't happy about **it and i** was in the middle of
and we share some other stuff about what our lives are like things that i can't ever tell people at home this girl i can tell her the really ugly stuff and she	have to go **home and i** went to see **her and she** said you know **what i** think **about** you **and i** know	a presentation with **some people** who were coming up for their first visit to a new place **i** had already told **her i** wanted to be friends but **she** seemed to

still understands how it can still be pretty she understands like how nice he's gonna be when i get home and how sweet that'll be	you're not **gonna be** able to help you with that but **i** don't know what to do **and i** just couldn't do **it and**	not feel the same way about **it** the thing is **when** you find out you're not good enough for someone to love you **and**
we are chain smoking off each other oh that's almost out come on and we we go through this entire pack until it's gone and then i say you know what uh this is a little funny but you're gonna	i'm like **oh i** don't **know what** i'm **gonna** do **is** i'll be able to get **out** of the car **and then i** got to the end of	**you** just have to fight the urge to throw the ring back in the box **and** take your pick of the two **and** set them free **but i** can't **i** need
have to show me the way to get home because although i'm twenty three years old i don't have my driver's license yet and i just jumped out right when i needed to and she says well why don't you come back to my house and i'll give you	**my** life so **i** go **to the** store **and i** went **to** college **and i** got a job at **the** university **and i** was like oh **my** god **i have to** go **and**	**to** ask **you** something about **my** childhood **i** had **to get my license to** drive so **my** mom could **have me out** of **the house and she** wanted **me to** go with her **to** visit **my** dad
a ride i say ok great and we start walking and uh we get to this um lots of uh lights and uh the roads are getting	**i** said okay **and** he **says** well you know **i** don't really want **to** do it but **i** had **to**	so **we** left **and** got back home **and** it was then that **i** noticed something was different about **the** way **the light** was
wider and wider and there's more cars and i see um lots of stores you know laundromats and dollar stores and emergecenters and	get out **of** the **car and i** got out **of** the **car and i** got out **of** the **car and i** got	changing around me **and i** looked up **and** saw a tall building with a few towers **and** a white flag flying above it **and** then
then we cross over us one and uh she leads me to some place and i think no but yes carl's efficiency	in the car **and i** was like oh my god this is the first time i've ever seen you **and i** was like well **i** don't know what **to** do	**we** came across a town about miles from our house **and i** thought oh god it was my grandmother the
apartments this girl lives there and it's horrible and it's lit up so bright just to illuminate the horribleness of it it's the kind of place where you drive your car right up and the door's	with **it and** i think **it's** a good thing i think **it's** a good thing i guess is that i'm a big fan **of the** old	old lady from **the** church that was **so** very special **to** us but **it** was very sad because **it** meant that she had a death certificate she was
right there and there's fifty million cigarette butts outside and there's like doors one through seven and you know behind every single door there's some horrible misery going on there's someone crying or drunk	black **and** white tv set in the living room of my parents house **and** i was the only **one** in my	**on** the front page of **every** newspaper **and** it was just the top of the times which meant that **there** were stories of her being beaten at her job **and** having a
or lonely or cruel and i think oh god she lives here how awful we go to the door door number four and she very very quietly	family **i** was **the** only girl in **the** whole world **and i** was like well **i** don't know what i'm gonna do is **i go to the** hospital **and**	boyfriend who beat her up but he wasn't that bad **to** begin with so **i** went **to the** cops **i** had **to** walk out of there **and**

keys in as soon as the door opens i hear the blare of television come out and on the blue light of the television the smoke of a hundred cigarettes in that little crack of light	**i** went back to **the** car **and i** got **out of the** car **and** into **the** back seat **of the** car **i** was sitting **in the** back **of** my head **and**	take my chances **and i** got into **a** car **that** was sitting there **the** passenger seat was full **of** beer bottles **and the** front windshield had glass **in** it **and**
and i hear the man and he says where were you and she says never mind i'm back and he says you alright and she says yeah i'm alright	**i'm** like what **the** fuck is wrong with **you and i'm** like well **i** don't know what to **say** so **i** tell **you** what **i'm** gonna	my mother was screaming at my father **and saying** oh no **you** aren't leaving us **he** said well why not **i** looked
and then she turns to me and says you want a beer and he says who the fuck is that and she pulls me over and he sees me and he says oh hey i'm not a threat	**say is** it okay if i go out with him **and** i said well **you** know **i'm not** sure i don't know what **to**	at my dad **who** was smiling i told my mom **to** call him back **and** when **he** finally answered **she** said no thanks my dad
just then he takes a drag of his cigarette a very hard hard drag you know the kind that makes the end of it really heat up hot hot hot and long and it's a little scary and	do so i went into **the** bathroom **and** i was in there for **a** while **and then** i would go to **the** hospital **and** get **the** best care i could get **and** i had to **take**	had some sort **of** breakdown **and just** threw **up** for days **it** was **a very** serious form **of** depression **that** was going around **and** i had to put
i follow the cigarette down because i'm afraid of that head falling off and i'm surprised when i see in the crook of his arm a little boy sleeping	**off** my shoes **and i** put them on **the** floor **and i** was like oh my god **i'm** so glad **i** didn't have to	my hands up to hide **the** fact **that** it hurt to have it pulled up **the** skin on **the** side **of** my face **and** over my eye
a toddler and i think and just then the girl reaches underneath the bed and takes out a carton and she taps out the last s pack of cigarettes in there	worry about my dad **and i** was like oh my god i'm **in the** shower **and i** get **in the** car **and** we drove **out** to **the**	my dad said what is that he walked up to my father grabbed **the** shotgun from his side **and** started loading it with
and on the way up she kisses the little boy and then she kisses the man and the man says again you alright and she says yeah i'm just gonna go out and smoke with her and so	beach **and** there's this guy **and** he's like oh my god i can't believe you're **going** to be	rounds i immediately pulled my brother to me **and** hugged him i **then** asked my sister if **she** could stay **with** us we all left to **go** get our cars **and**
we go outside and sit amongst the cigarette butts and smoke and i say wow that's your little boy and she says yeah isn't he beautiful and i say yeah he is he is beautiful	a couple of days later **i** get a call from my dad **and i** was like well i'm not sure **i** don't know what to **say** but **i**	drove to **the** beach at am on **the** way back **he** said **i** can't believe this **i** don't understand it i'm like oh man you must be
he's my light he keeps me going she says we finish our cigarettes she finishes her beer i don't have a beer cause i can't go home	know you're not **going** to like this but **i** think it was the first time **i** had sex with you and **i** think you're	really special if you're the best thing in **my** life and **i** can barely speak **i** will probably never **have** sex again for fear of

with beer on my breath and she goes inside to get the keys she takes too long in there getting the keys and i	gonna have to do a lot of work on the car and i got out of the car and	becoming addicted to the stuff and then it gets worse when she comes out of it i start to worry that something is
think something must be wrong and she comes out and she says look i'm really sorry but um like we don't have any gas in the car it's already on e and he needs to get to work in the morning and um	i got out of the car and i went to the bathroom and i went in there and i was like oh my god i can't believe i'm having a baby and i was like	wrong but then she says to me hey do you think she can afford to pay rent on time to a school for her kids to go to and
i you know i i'm gonna be walk to work as it is so what i did was though here look i drew out this map for you and you're really you're like a mile and a half from home and um if you walk three streets over you'll be back on that pretty street	well i don't know what to do so i told him that i was going to do it and then i got a call from one of the guys who had been in the hospital for a	she would just say oh sorry i did the math on this one i will send her a link to a different site in the same time period to see how far back
and you just take that and you'll be fine and she also has wrapped up in toilet paper seven cigarettes for me a third of her pack i note and a new pack of	couple of days and then i had to put on a pair of tiny white leering eyes and my hair was done and i had to get my	she gets so i do this too i put a bunch of photos in an envelope with a few other stuff i brought a box of
matches and she tells me good bye and that was great to meet you and how lucky and that was fun and you know let's be friends	hair done and then i go home and i tell my mom i love her so much but i just couldn't do it	them and he thanked me again then we started to talk about our life and the family he was from and i
and i say yeah ok and i walk away but i kind of know we're not gonna be friends i might not ever see her again and i kind of	and i'm like well i don't know what to do so i go to university and i got a job at the	said yes and then i got angry when i thought about it because i wasn't sure what was happening i didn't really care at all
know i don't think she's ever going to be a vet and i cross and i walk away and maybe this would've seemed like a visit from my possible future and scary but it	university and i get a job at the club and then i went back to my room and i sat in my room for hours on end and i was	until i saw it was the girl from my dreams that was driving down the road and i started to notice that the car was a bit older than my own i couldn't
kind of does the opposite on the walk home i'm like man that was really grim over there and i'm going home now to my nice boyfriend	like oh my god i'm so sorry i don't know what i was going to do a lot of things that i didn't want to be	make it out clearly so i yelled hey you have to stop i just saw a man that i knew in the past who was very physically

and he is gonna be so extra happy to see me and we have a one bedroom apartment and we have two trees and there's a yard and we have this jar in	**a** girl **and** i was **so happy and** i was **so** excited **and** i was like oh my god **this is** the first time i've	abused by his wife she had him committed **and** was now living **in** his **bedroom** where **he** had **a** couch **and** tv the **apartment** was about feet long **and** had
the kitchen where there's like loose money that we can use for anything like we would never ever run out of gas and um i don't have a baby you know so i can leave whenever i want	**ever** been to **and i** was **like** oh my god **i** can't believe you're	two bathrooms **and a** huge closet to store **the** clothes **i** didn't **have** any **money** or food because **i** wasn't hungry **i** was lucky my parents didn't lock me up **so**
i smoked all seven cigarettes on the way home and people who have never smoked cigarettes just think ick disgusting and poison	not going to see him again **the** next day **and** you know **i think** it's a really good question **and i** don't know what to do	**i just** started drinking **all the** time **the** day before thanksgiving **and** my mom told me that **the** reason she doesn't like alcohol is because it's addictive **i** don't

Table S11. Performance of language generation without text prompt (averaged across participants) in different datasets. The comparison between BrainLLM and PerBrainLLM are significant at $q(FDR) < 0.05$ (one-sided non-parametric test) on all datasets and metrics, respectively.

Dataset	Model	BLEU-1(↑)	ROUGE-1(↑)	ROUGE-L(↑)	WER(↓)	Win rate (with PerBrainLLM)
Pereira's	PerBrainLLM	0.0787	0.0553	0.0540	0.9726	0.5000
	BrainLLM	**0.1025**	**0.0788**	**0.0749**	**0.9610**	**0.8885**
Huth's	PerBrainLLM	0.0960	0.0817	0.0779	0.9703	0.5000
	BrainLLM	**0.1356**	**0.1160**	**0.1099**	**0.9541**	**0.8816**
Narratives	PerBrainLLM	0.1270	0.1133	0.1092	0.9328	0.5000
	BrainLLM	**0.1320**	**0.1184**	**0.1145**	**0.9283**	**0.6728**

Table S12. Performance of language generation with LLM with different sizes of parameters in different datasets (averaged across participants). As we focus on the performance comparison between BrainLLM and PerBrainLLM, we did not show experiments with StdLLM here. But you can find more results on StdLLM in our github repository (https://github.com/YeZiyi1998/Brain-language-generation). $*$ denotes a significant difference with BrainLLM using a Wilcoxon test with $q(FDR) < 0.5$ under the same model and the same dataset.

Dataset	LLM backbone	Model	BLEU-1(↑)	ROUGE-1(↑)	ROUGE-L(↑)	WER(↓)
Pereira's	Llama-2 (7B)	StdLLM	0.2415*	0.2133*	0.2096*	0.8349*
		PerBrainLLM	0.3269*	0.2815*	0.2751*	0.7783*
		BrainLLM	**0.3432**	**0.2987**	**0.2878**	**0.7576**
	GPT-2-xl (1.5B)	PerBrainLLM	0.2772	0.234	0.2256	0.8246
		BrainLLM	0.2814*	0.2378*	0.2292*	0.8239*
	GPT-2-large (774M)	PerBrainLLM	0.2605*	0.213*	0.2057*	0.8404*
		BrainLLM	0.2655	0.2182	0.2106	0.8395
	GPT-2-medium (345M)	PerBrainLLM	0.2100	0.1649*	0.1605	0.8774
		BrainLLM	0.2118	0.1672	0.1626	0.8779
	GPT-2 (117M)	PerBrainLLM	0.1866	0.1456	0.1426	0.8968
		BrainLLM	0.1846	0.1445	0.1414	0.8973
Huth's	Llama-2 (7B)	StdLLM	0.1500*	0.1360*	0.1310*	0.92*
		PerBrainLLM	0.1668	0.1536	0.1474	0.9109
		BrainLLM	**0.1899**	**0.1780**	**0.1709**	**0.8916**
	GPT-2-xl (1.5B)	PerBrainLLM	0.1708*	0.1652*	0.1581*	0.909 *
		BrainLLM	0.1791	0.1729	0.1656	0.9022
	GPT-2-large (774M)	PerBrainLLM	0.1657*	0.1584*	0.1516*	0.9132*
		BrainLLM	0.1762	0.1693	0.1616	0.9049
	GPT-2-medium (345M)	PerBrainLLM	0.164*	0.1549*	0.1489*	0.914*
		BrainLLM	0.1667	0.1578	0.1514	0.9126
	GPT-2 (117M)	PerBrainLLM	0.1088*	0.1059*	0.0997*	0.9516*
		BrainLLM	0.1096	0.1065	0.1011	0.952
Narratives	Llama-2 (7B)	StdLLM	0.0953*	0.0858*	0.0829*	0.9485*
		PerBrainLLM	0.1269*	0.1211*	0.1105*	0.9311*
		BrainLLM	**0.1375**	**0.1301**	**0.1209**	**0.9239**
	GPT-2-xl (1.5B)	PerBrainLLM	0.1248*	0.1171*	0.1121*	0.9340*
		BrainLLM	0.1298	0.122	0.1168	0.9319
	GPT-2-large (774M)	PerBrainLLM	0.1202*	0.1124*	0.1074*	0.9402*
		BrainLLM	0.1237	0.1159	0.1104	0.9401
	GPT-2-medium (345M)	PerBrainLLM	0.1056*	0.0993*	0.095*	0.9472*
		BrainLLM	0.1063	0.0999	0.0956	0.9463
	GPT-2 (117M)	PerBrainLLM	0.1099*	0.1032*	0.098*	0.9509*
		BrainLLM	0.1111	0.1047	0.0997	0.9506

Table S13. Comparison of language generation performance (averaged across participants) of BrainLLM and the pre-construction followed by post-hoc selection model (Tang et al. [58]) in Huth's dataset. $\dagger/*$ denotes a significant difference with BrainLLM/PerBrainLLM using a Wilcoxon test with $q(FDR) < 0.5$ under the setting (with or without text prompt). The win rate for the Tang et al. [58]'s model is not available as the selection-based method can not get the possibilities of generating the perceived continuation. The original work proposed by Tang et al. [58] utilizes settings similar to generation without any text prompts. Hence, we present their performance comparison in both settings.

Setting	Model	BLEU-1(↑)	ROUGE-1(↑)	ROUGE-L(↑)	WER(↓)	Win rate (with PerBrainLLM)
with text prompt	PerBrainLLM	0.1668*	0.1536*	0.1474*	0.9200*	0.5000*
	Tang et al. [58]	0.1675*	0.1537*	0.1483*	0.9197*	-
	BrainLLM	**0.1899**†	**0.178**†	**0.1709**†	**0.8916**†	**0.7667**†
without text prompt	PerBrainLLM	0.0960*	0.0817*	0.0779*	0.9703*	0.5000*
	Tang et al. [58]	0.0967†,*	0.0818*	0.0788†,*	0.9700†,*	-
	BrainLLM	**0.1356**†	**0.1160**†	**0.1099**†	**0.9541**†	**0.8816**†

Table S14. Statistics of the LLMs adopted in our experiments. These statistics are gathered from the original paper [61, 51] and the public sourced repositories (https://huggingface.co/meta-llama/Llama-2-7b and https://huggingface.co/gpt2).

Model	#Parameters	#Transformer layers	Embedding size	Vocabulary size	Quantization	#Max input tokens
Llama-2	7B	32	4,096	32,000	float16	4,096
GPT-2-xl	1.5B	48	1,600	50,257	float32	1,024
GPT-2-large	774M	36	1,280	50,257	float32	1,024
GPT-2-medium	345M	24	1,024	50,257	float32	1,024
GPT-2	117M	12	768	50,257	float32	1,024

Table S15. Overall statistics of fMRI datasets.

Dataset	Stimuli	#Participants	#Total Duration	#Duration per participant	#Total TRs	#TRs per participant	#Total words	#Words per participant
Pereira's	visual	5	7.0 h	1.4 h	3135	627	38650	7730
Huth's	auditory	8	3.5 days	10 h	122992	15374	427296	53412
Narratives	auditory	28	21.0h	45 min	48496	1732	230460	8231

Table S16. Language generation performance averaged across participants in different datasets. The difference between BrainLLM and PerBrainLLM/StdLLM are significant at $q(FDR) < 0.05$ (one-sided non-parametric test) on all datasets and metrics, respectively.

Dataset	Model	BLEU-1(↑)	ROUGE-1(↑)	ROUGE-L(↑)	WER(↓)
Pereira's	StdLLM	0.2415	0.2133	0.2096	0.8349
	PerBrainLLM	0.3269	0.2815	0.2751	0.7783
	BrainLLM	**0.3432**	**0.2987**	**0.2878**	**0.7576**
Huth's	StdLLM	0.1500	0.1360	0.1310	0.9200
	PerBrainLLM	0.1668	0.1536	0.1474	0.9109
	BrainLLM	**0.1899**	**0.1780**	**0.1709**	**0.8916**
Narratives	StdLLM	0.0953	0.0858	0.0829	0.9485
	PerBrainLLM	0.1269	0.1211	0.1105	0.9311
	BrainLLM	**0.1375**	**0.1301**	**0.1209**	**0.9239**

Table S17. Win rate of BrainLLM versus PerBrainLLM across different selections of the Brain adapter structures in Huth's dataset.

Structure	MLP+pos	MLP	concat+MLP	RNN	Linear
Win rate	**0.7893**	0.7782	0.6826	0.6755	0.7155

Table S18. Language similarity metrics of BrainLLM with different training protocols with the backbone LLM GPT2-xl in Huth's dataset.

Training protocol	BLEU-1	ROUGE-1	ROUGE-L	WER
Full parameter fine-tuning	0.1331	0.1228	0.1180	0.9342
Combined	0.1672	0.1611	0.1523	0.9197
Prompt-tunning	**0.1791**	**0.1729**	**0.1656**	**0.9022**

Table S19. Win rate of BrainLLM versus PerBrainLLM across different selections of the learning rate and reduction dimensionality in Huth's dataset. We set the learning rate to 1×10^{-4} and choose the reduction dimensionality c as 1,000 which offers the best performance while balancing computational efficiency. We then consistently apply these parameters across all datasets and in all further analyses.

Learning rate	$1e^{-3}$	$1e^{-4}$	$1e^{-5}$	$1e^{-6}$
Win rate	0.7334	**0.7893**	0.7746	0.7240
Reduction dimensionality	500	1,000	2,000	4,000
Win rate	0.7532	**0.7893**	0.7813	0.7767

Table S20. Experimental results of the full-text construction task for a perceived story "where there's smoke" with a range of language similarity metrics in three subjects of Huth's dataset. (1) A floor for each metric was computed by scoring the mean similarity between the actual stimulus words and a sequence generated from a language model without using any brain data ("null"). Here Tang et al. [58] uses a private language model which is trained in a corpus constructed with Reddit stories and is similar to the perceived story content. On the other hand, we use a publicly available language model GPT2-xl which is trained in a general corpus and therefore shows worse performance when compared to Tang et al. [58] when no brain input is given. (2) When given brain input, the proposed BrainLLM outperforms Tang et al. [58] in all language similarity metrics. A pair-wise t-test shows significant improvements are observed particularly in METEOR ($p = 3e^{-5}$), and in WER ($p = 0.03$). (3) We also test a more fair control with the brain input permutated. The experimental results demonstrate that our method significantly outperforms the control model in metrics of BLEU-1 ($p = 3e^{-4}$), and WER ($p=5e^{-4}$). (4) A ceiling for each metric was computed by manually translating the actual stimulus words into Mandarin Chinese, and automatically translating the words back into English using a state-of-the-art machine translation system. The numbers are inherited from Tang et al. [58]'s paper.

Input	**Method**	**Metrics**		
		BLEU-1(↑)	**WER** (↓)	**METEOR**(↑)
Null (without brain input)	Tang et al. [58]	0.1908	0.9637	0.1323
	GPT2-xl	0.1417	0.9569	0.1181
Permutated brain input of Subject 1	Tang et al. [58]	0.1967	0.9703	0.1542
	BrainLLM	0.2363	0.9499	0.1995
Brain input of Subject 1	Tang et al. [58]	0.2331	0.9407	0.1621
	BrainLLM	**0.2539**	**0.9158**	**0.2078**
Permutated brain input of Subject 2	Tang et al. [58]	0.2016	0.9761	0.1586
	BrainLLM	0.2161	0.9545	0.1988
Brain input of Subject 2	Tang et al. [58]	0.2426	0.9354	0.1677
	BrainLLM	**0.2518**	**0.9259**	**0.2031**
Permutated brain input of Subject 3	Tang et al. [58]	0.2173	0.9576	0.1695
	BrainLLM	0.2068	0.9755	0.2011
Brain input of Subject 3	Tang et al. [58]	0.247	0.9243	0.1703
	BrainLLM	**0.2497**	**0.9190**	**0.218**
Mandarin	machine translation	0.4363	0.7459	0.3991

References

[1] Hervé Abdi and Lynne J Williams. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459, 2010.

[2] Nicolas Affolter, Beni Egressy, Damian Pascual, and Roger Wattenhofer. Brain2word: decoding brain activity for language generation. *arXiv preprint arXiv:2009.04765*, 2020.

[3] Andrew James Anderson, Douwe Kiela, Jeffrey R Binder, Leonardo Fernandino, Colin J Humphries, Lisa L Conant, Rajeev DS Raizada, Scott Grimm, and Edmund C Lalor. Deep artificial neural networks reveal a distributed cortical network encoding propositional sentence-level meaning. *Journal of Neuroscience*, 41(18):4100–4119, 2021.

[4] Richard Antonello, Aditya Vaidya, and Alexander G Huth. Scaling laws for language encoding models in fmri. *arXiv preprint arXiv:2305.11863*, 2023.

[5] Gopala K Anumanchipalli, Josh Chartier, and Edward F Chang. Speech synthesis from neural decoding of spoken sentences. *Nature*, 568(7753): 493–498, 2019.

[6] Khai Loong Aw and Mariya Toneva. Training language models to summarize narratives improves brain alignment. In *Eleventh International Conference on Learning Representations*. OpenReview. net, 2023.

[7] Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, et al. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*, 2022.

[8] Jeffrey R Binder and Rutvik H Desai. The neurobiology of semantic memory. *Trends in cognitive sciences*, 15(11):527–536, 2011.

[9] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[10] Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior research methods*, 46:904–911, 2014.

[11] Charlotte Caucheteux and Jean-Rémi King. Brains and algorithms partially converge in natural language processing. *Communications biology*, 5(1):134, 2022.

[12] Charlotte Caucheteux, Alexandre Gramfort, and Jean-Rémi King. Deep language algorithms predict semantic comprehension from brain activity. *Scientific reports*, 12(1):16327, 2022.

[13] Shweta Chauhan and Philemon Daniel. A comprehensive survey on various fully automatic machine translation evaluation metrics. *Neural Processing Letters*, pages 1–55, 2022.

[14] Michael WL Chee, Chun Siong Soon, Hwee Ling Lee, and Christophe Pallier. Left insula activation: a marker for language attainment in bilinguals. *Proceedings of the National Academy of Sciences*, 101(42):15265–15270, 2004.

[15] Junhyeong Cho, Gilhyun Nam, Sungyeon Kim, Hunmin Yang, and Suha Kwak. Promptstyler: Prompt-driven style generation for source-free domain generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15702–15712, 2023.

[16] Andy Clark. Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, 36(3): 181–204, 2013.

[17] Alexandre Défossez, Charlotte Caucheteux, Jérémy Rapin, Ori Kabeli, and Jean-Rémi King. Decoding speech perception from non-invasive brain recordings. *Nature Machine Intelligence*, pages 1–11, 2023.

[18] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202, 1980.

[19] John DE Gabrieli, Russell A Poldrack, and John E Desmond. The role of left prefrontal cortex in language and memory. *Proceedings of the national Academy of Sciences*, 95(3):906–913, 1998.

[20] Deep Ganguli, Danny Hernandez, Liane Lovitt, Amanda Askell, Yuntao Bai, Anna Chen, Tom Conerly, Nova Dassarma, Dawn Drain, Nelson Elhage, et al. Predictability and surprise in large generative models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 1747–1764, 2022.

[21] Ariel Goldstein, Zaid Zada, Eliav Buchnik, Mariano Schain, Amy Price, Bobbi Aubrey, Samuel A Nastase, Amir Feder, Dotan Emanuel, Alon Cohen, et al. Shared computational principles for language processing in humans and deep language models. *Nature neuroscience*, 25(3): 369–380, 2022.

[22] John T Hale, Luca Campanelli, Jixing Li, Shohini Bhattasali, Christophe Pallier, and Jonathan R Brennan. Neurocomputational models of language processing. *Annual Review of Linguistics*, 8:427–446, 2022.

[23] Shilin He, Jieming Zhu, Pinjia He, and Michael R Lyu. Loghub: A large collection of system log datasets towards automated log analytics. *arXiv preprint arXiv:2008.06448*, 2020.

[24] Shaohan Huang, Li Dong, Wenhui Wang, Yaru Hao, Saksham Singhal, Shuming Ma, Tengchao Lv, Lei Cui, Owais Khan Mohammed, Barun Patra, et al. Language is not all you need: Aligning perception with language models. *Advances in Neural Information Processing Systems*, 36, 2024.

[25] Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*, 2020.

[26] Timothy A Keller, Patricia A Carpenter, and Marcel Adam Just. The neural bases of sentence comprehension: a fmri examination of syntactic and lexical processing. *Cerebral cortex*, 11(3): 223–237, 2001.

[27] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[28] Sasa L Kivisaari, Marijn van Vliet, Annika Hultén, Tiina Lindh-Knuutila, Ali Faisal, and Riitta Salmelin. Reconstructing meaning from bits of information. *Nature communications*, 10 (1):927, 2019.

[29] Dietrich Klakow and Jochen Peters. Testing the correlation of word error rate and perplexity. *Speech Communication*, 38(1-2):19–28, 2002.

[30] Marta Kutas and Steven A Hillyard. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427):203–205, 1980.

[31] Amanda LeBel, Lauren Wagner, Shailee Jain, Aneesh Adhikari-Desai, Bhavin Gupta, Allyson Morgenthal, Jerry Tang, Lixiang Xu, and Alexander G Huth. A natural language fmri dataset for voxelwise encoding models. *Scientific Data*, 10(1):555, 2023.

[32] Yulia Lerner, Christopher J Honey, Lauren J Silbert, and Uri Hasson. Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *Journal of Neuroscience*, 31(8): 2906–2915, 2011.

[33] Brian Lester, Rami Al-Rfou, and Noah Constant. The power of scale for parameter-efficient prompt tuning. *arXiv preprint arXiv:2104.08691*, 2021.

[34] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55 (9):1–35, 2023.

[35] Xiao Liu, Kaixuan Ji, Yicheng Fu, Weng Tam, Zhengxiao Du, Zhilin Yang, and Jie Tang. P-tuning: Prompt tuning can be comparable to fine-tuning across scales and tasks. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 61–68, 2022.

[36] Xiao Liu, Yanan Zheng, Zhengxiao Du, Ming Ding, Yujie Qian, Zhilin Yang, and Jie Tang. Gpt understands, too. *AI Open*, 2023.

[37] Yifei Luo, Minghui Xu, and Deyi Xiong. Cogtaskonomy: Cognitively inspired task taxonomy is beneficial to transfer learning in nlp. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 904–920, 2022.

[38] Gary Lupyan and Andy Clark. Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science*, 24(4):279–284, 2015.

[39] Giulio Mecacci and Pim Haselager. Identifying criteria for the evaluation of the implications of brain reading for mental privacy. *Science and Engineering Ethics*, 25:443–461, 2019.

[40] Clara Meister and Ryan Cotterell. Language model evaluation beyond perplexity. *arXiv preprint arXiv:2106.00085*, 2021.

[41] Tom M Mitchell, Svetlana V Shinkareva, Andrew Carlson, Kai-Min Chang, Vicente L Malave, Robert A Mason, and Marcel Adam Just. Predicting human brain activity associated

with the meanings of nouns. *science*, 320(5880): 1191–1195, 2008.

[42] David A Moses, Sean L Metzger, Jessie R Liu, Gopala K Anumanchipalli, Joseph G Makin, Pengfei F Sun, Josh Chartier, Maximilian E Dougherty, Patricia M Liu, Gary M Abrams, et al. Neuroprosthesis for decoding speech in a paralyzed person with anarthria. *New England Journal of Medicine*, 385(3):217–227, 2021.

[43] Mariacristina Musso, Andrea Moro, Volkmar Glauche, Michel Rijntjes, Jürgen Reichenbach, Christian Büchel, and Cornelius Weiller. Broca's area and the language instinct. *Nature neuroscience*, 6(7):774–781, 2003.

[44] Samuel A Nastase, Yun-Fei Liu, Hanna Hillman, Asieh Zadbood, Liat Hasenfratz, Neggin Keshavarzian, Janice Chen, Christopher J Honey, Yaara Yeshurun, Mor Regev, et al. The "narratives" fmri dataset for evaluating models of naturalistic language comprehension. *Scientific data*, 8(1):250, 2021.

[45] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744, 2022.

[46] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318, 2002.

[47] Xiaomei Pei, Dennis L Barbour, Eric C Leuthardt, and Gerwin Schalk. Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. *Journal of neural engineering*, 8(4): 046028, 2011.

[48] Francisco Pereira, Bin Lou, Brianna Pritchett, Samuel Ritter, Samuel J Gershman, Nancy Kanwisher, Matthew Botvinick, and Evelina Fedorenko. Toward a universal decoder of linguistic meaning from brain activation. *Nature communications*, 9(1):963, 2018.

[49] Amy R Price, Michael F Bonner, Jonathan E Peelle, and Murray Grossman. Converging evidence for the neuroanatomic basis of combinatorial semantics in the angular gyrus. *Journal of Neuroscience*, 35(7):3276–3284, 2015.

[50] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. OpenAI. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pd 2018.

[51] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

[52] Stephen Rainey, Stéphanie Martin, Andy Christen, Pierre Mégevand, and Eric Fourneret. Brain recording, mind-reading, and neurotechnology: ethical issues from consumer devices to brain-based speech decoding. *Science and engineering ethics*, 26:2295–2311, 2020.

[53] Riitta Salmelin, A Schnitzler, L Parkkonen, K Biermann, Päivi Helenius, K Kiviniemi, K Kuukka, F Schmitz, and H-J Freund. Native language, gender, and functional organization of the auditory cortex. *Proceedings of the National Academy of Sciences*, 96(18):10460–10465, 1999.

[54] Martin Schrimpf, Idan Asher Blank, Greta Tuckute, Carina Kauf, Eghbal A Hosseini, Nancy Kanwisher, Joshua B Tenenbaum, and Evelina Fedorenko. The neural architecture of language: Integrative modeling converges on predictive processing. *Proceedings of the National Academy of Sciences*, 118(45): e2105646118, 2021.

[55] Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul F Christiano. Learning to summarize with human feedback. *Advances in Neural Information Processing Systems*, 33:3008–3021, 2020.

[56] Jingyuan Sun, Shaonan Wang, Jiajun Zhang, and Chengqing Zong. Towards sentence-level brain decoding with distributed representations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 7047–7054, 2019.

[57] Jingyuan Sun, Shaonan Wang, Jiajun Zhang, and Chengqing Zong. Neural encoding and decoding with distributed sentence representations. *IEEE Transactions on Neural Networks and Learning Systems*, 32(2):589–603, 2020.

[58] Jerry Tang, Amanda LeBel, Shailee Jain, and Alexander G Huth. Semantic reconstruction of continuous language from non-invasive brain

recordings. *Nature Neuroscience*, pages 1–9, 2023.

[59] Mariya Toneva. *Bridging Language in Machines with Language in the Brain*. PhD thesis, Carnegie Mellon University, 2021.

[60] Mariya Toneva and Leila Wehbe. Interpreting and improving natural-language processing (in machines) with natural language-processing (in the brain). *Advances in neural information processing systems*, 32, 2019.

[61] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.

[62] Helene Van Ettinger-Veenstra, Anita McAllister, Peter Lundberg, Thomas Karlsson, and Maria Engström. Higher language ability is related to angular gyrus activation increase during semantic processing, independent of sentence incongruency. *Frontiers in human neuroscience*, 10:110, 2016.

[63] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[64] Francis R Willett, Donald T Avansino, Leigh R Hochberg, Jaimie M Henderson, and Krishna V Shenoy. High-performance brain-to-text communication via handwriting. *Nature*, 593(7858): 249–254, 2021.

[65] Nuwa Xi, Sendong Zhao, Haochun Wang, Chi Liu, Bing Qin, and Ting Liu. Unicorn: Unified cognitive signal reconstruction bridging cognitive signals and human language. *arXiv preprint arXiv:2307.05355*, 2023.

[66] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*, 2019.

[67] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. A survey of large language models. *arXiv preprint arXiv:2303.18223*, 2023.

[68] Danhua Zhu, Jordi Bieger, Gary Garcia Molina, and Ronald M Aarts. A survey of stimulation methods used in ssvep-based bcis. *Computational intelligence and neuroscience*, 2010, 2010.

[69] Shuxian Zou, Shaonan Wang, Jiajun Zhang, and Chengqing Zong. Towards brain-to-text generation: Neural decoding with pre-trained encoder-decoder models. In *NeurIPS 2021 AI for Science Workshop*, 2021.