Profits programmed and the control of the control o	La real pile
Notes to state of the pattern of the	
The main file of the state of t	
The lates are the series of the control of the cont	
The logic of the control of the cont	
Training feet that these little demonstrates are present and the state of the state	
# Silent train away postering to make some a same prior - protect examing to make some a same prior - come is the man managing and makes some a same prior - come is the man managing and makes some a same prior - come is the man managing and makes some a same prior - come is the man managing and makes some a same prior - come is the man managing and makes some a same prior - come is the man managing and makes some a same prior - come is the man managing and makes some a same prior - come is the man managing and makes some a same prior - come is the man managing and makes some a same prior - come is the man managing and makes some a same prior - come is the managing and makes some a same prior - come is the managing and makes some a same prior - come is the managing and makes some a same prior - come is the managing and makes some a same prior - come is the managing and makes some a same prior - come is the managing and makes some a same prior - come is the managing and makes some a same prior - come is the managing and makes some a same prior - come is the managing and makes some prior - come is the managing and makes some prior - come is the managing and makes some prior - come is the managing and makes some prior - come is the managing and makes some prior - come is the managing and makes some prior - come is the managing and makes some prior - come is the managing and makes some prior - come is the managing and makes some prior - come is the managing and makes some prior - come is the managing and makes - come i	
Printers extending 14 mat in these sums at time grants The man is a man in manager and manager and manager at the printers of the manager and manager	
prolifered externing 1s with in these sums at least professions and least professions and least professions are least and the profession and the p	
The set of the control of the second states and a second of alternating six and acting second acting	
3. Maked antended to decessioning transformers, tot amonthm is experient cancil marked), making a trace can only attend to provide tension in the Equinitic, not deliver could, their the model agreement serves in an enter-regression makes Detailed composition freqs . Can tested embased varior to transformer and their vectors; during (b), pag (R), and value (v). Three values are set to determine the aprication leaves . Can tested embased varior to transformer and their vectors; during (b), pag (R), and value (v). Three values are set to determine the aprication leaves . Can tested a section of the area of the country of the c	
Detailed composition strongs Consider composition strongs Consid	
Deballed composition Steps 1. Each status companies of the formal into 3 different vectors; during (0), Reg (R), and value (v). There years are seed to dedermine the approximation forms 1. Each status control of the formal into 3 different vectors; during (0), Reg (R), and value (v). There years are years to dedermine the approximation forms 1. Each status control of the formal into 3 different vectors; 1. Each status 1. V = X Nu ^N 1. V = A nu ^N 1. Sit appearance of the formal vectors are interested and a status formal into the 3 vectors above, below is when they signify; 1. Guerg (0): a query 1 = a teprospectation of a space when do store two well to majority every other token, Essentially when analysis are formal into the general status of the control of the c	
I. Each tables combissed valvor is transformed into 3 different vectors; Query (2), Reg (K), and value (V), Thrist values are used to dedicate the appendix the appendix of the second to the second t	
I. Each tables combissed valvor is transformed into 3 different vectors; Query (2), Reg (K), and value (V), Thrist values are used to dedicate the appendix the appendix of the second to the second t	
I. Each tables combissed valvor is transformed into 3 different vectors; Query (2), Reg (K), and value (V), Thrist values are used to dedicate the appendix the appendix of the second to the second t	
- A lavery = X W - X - X - X - X - X - X - X - X - X	
**K = XWK **V = XNV **PECK WA WE, and M" are parameter indirects that are located during scaling why brook up late 3 different vectors **So and Self-attention method also of a frankformer, each laport vector is evans formed into the 3 vectors above, below is what they signify: 1 during (a): a query is a representation of a force vised to store two west to material every office token, Essentially when considering a specific token, it query seek out which tokens (Ruys) are most relevant to it. 1. Key (N): A Key is associated with a token to be majored abovened abovened abovened a query is compared to this Key; the resulting score determines the impact of the years of the parameters are also above.	
tere was the assertion methodises there are leasted during training "The section methodism of a transformer, each input vector to transformed into the 3 vectors above, below is what they signify: I durry (a): a query is a representation of a token used to Score town was it may make your token. Essentially when considering a specific token, it query seek out which tokens (Right) are more relevant to it. 1. Key (K): A Key is associated with a token to be majored against query; before the token to this key; the resulting score decembers the impact as the value on the autopath.	
Here Way and Was are parameter regretes that are located during training why proces up late a different vectors "So and Self-attention mechanism of a frantformer, each input vector is transformed into the 3 vectors above, below is what they lightly: I divery (a): a query is a representation of a force used to Sepre vow well to majores every other token. Essentially when considering a Specific token, it query seek or which tokens (Rugh) are report relevant to its 1. Key (K): A Key is associated with a token to be majored against queries when a query is compared to this Key; the resulting secret determines the impact of the value on the autopate.	
why brook up have 3 different vectors "To not Self-attention methodismose, each hapot vector to evans formed into the 3 vectors above, below is what they signify: I durry (a): a query is a representation of a force weed to store two west to material every office token. Essentially when considering a specific token, it query seek out which tokens (regs) are most relevant to in I. Key (K): A Key is associated with a token to be majored against quertes hims a query is compared to this Key; the resulting score describing has the value on the output	
*So and Self-attention methanism of a transformer, each input vector is transformed into the 3 vectors above, below is what they lightly: 1 divery (a): a query is a representation of a force used to Sepre vow west to majorise every other token. Essenticilly when considering a Specific token, it query seek put which tokens (Ruys) are report relevant to its 1. Key K): A Key is associated with a token to be majorise applies given a query is compared to this Key; the resulting secre determines the impact of the value on the automate applies.	
Seek Dot which tokens (Rigs) are 1808- relevant to 19 1. Key (K): A Key 15 associated with a token to the majores abains quertes before a query is compared to this Key, the resulting score determines the impact of the	
2. Kcylk): A Kcy is associated with a boken to the majorice abalist quertes from a query is compared to this key; the resulting bear decimines the impact of the	uction is ased to
40/01 00 40C 014D07	
	corresponding toxen's
3. Values (4): A value is a representation of a sometiment to open the amount of the final output. Once tokens are search based on the query-key marries, the values	
	, are weighted by
faces figures and surject up to produce the output for the rest layer	
Assertion Scores determine from swin focus to place on other points of the input lequence when processing a specific part of that sequence.	
*A function scale calculation = Q RT Jak	
the scores are scaled down by the square root of the dimension of the key vectors (de) to Stabolize the gradients during training	
4. Once one model now processed the injurt sequence during the pre-fill strange are sequently sequence and the pre-fill strange are served. I strange primes the model with an initial test of sories the until one	
a) as hat for future generation	

		e																														
		examp		ue erra	ur re	ames	tre ver	sent f	e rest																							
		↓ €, €		E.		↓ F ₅ ,	ا د و	Ι Ι , ε,		€; →	word	enbid	tog +	p 051+1ana	, encode	ng for	word i															
								7 E					•																			
		-, ,	2 -	3 **	•	•• •	e, t	7 E	١																							
	ν,	c ga	ءَ له	40 Crc	mt c d	ncw	501 6	of emi	bcdd?no	gs wa	icre (each c	401 q	Constders	4nc W	ords a	10000 17 4	desc	MINE	115 m	ean) n	g Le	verg Ve	cso(S)								
	ro	COM	ote to	is 441	ic a	nc (du)	Vector	we	and M	10(4)014	ing te	by	tre e	mbedding	vcc tors																	
			1	the e	A+#15	05 41	חים פיות	atrtz	an po	na mch	es of	fec e	node i C	mcantag	drue be	haver	is Icaraca	from de	rter)													
			wą.	E, - '	₹, ≁ ‹	lnerd	A CCHOL									1	Ĕ,	,	,				5									
	40	rals c	OMPU	etton o	nell y	n gut	a 41	ery ve	wor fo	or eve	ry te	Men en	****	,			512		`	~a	512	=	Q,	! '								
																				64												
						٦	1 nerd a	1 ue crea	40rc r	J	7	l l	ores)	→																		
					W	E, ₩	E2	j wa	↓ ω	aj wa	wa Wa	J Wa	ļ	Con	rce phoa Il y	trus I	checking	is taci	, an	any w	10.43 f	nat c	tange to	t mpug	wranng							
	۰ س ٫	s 45	eg r	atrix																												
		a	→ Ē	• w _¢	к,	k,∙Q,	ĸ,·Q,	k, Q,	k,·Q,	K, · Q, 1	K, • Q,	K, · Q.	K, · Qa	٦		2 do	t product	Adm(7	hlqh	ighetel	with	≠nts	an high	Values								
	e	luffy	-→ E	ω¢	K2	k ₂ ∙ 0,	k,·Q,	K ₂ · Q ₃	k 2 0	k ₂ · Q ₆ I	ر. م. ا	k ₂ · Q,	k, Q,																			
		DIVIC	→ Ē.	ω,	κ,	k3. 0¹	k ₃ ∙Q	k,∙Q,	K3. G ^A	K3 • Q5 1	Kg · Q	K3 · Q7	K3 · Q1																			
	c	rcatu	e → Ē	• ω _ε	Кц	k". oʻ	k ₄ ·Q	K, · Q,	K4. G*	K4. 0* 1	K, · Q,	Kų· Q,	k4. G₽	7 6 11	*****	iet gr	oducts are	berng	6 m	attà p	ara)le	.3										
	rb	amed	→ €	₩.	K 5	Kg·Qį	K5. G	K ₅ ·Q ₃	K _S · Q ₁	K _E · Q _S	kg· Q	k _e ∙ Q,	Kg∙ Q ø																			
		:																														
• 60	nec pł	valle	4021	i 15	an swc	ring			≱ /4	the c	to 1	roduct	03 Q	Key a	nd quen	10010	f (1 feathy	high	tnen ·	inc sp	ecifre	key	embeddia	gs a44	ind to	4ne	dacua	embed	atngs			
				y wa																												
				i5 c	nangi	~9																										
	its	MC	rutug																													
2.	صالم ک						/ Dar	ς Δ.		.00	. 00			Earn		Java.	eath word	10.0														
																	atch coine								abii:tu	dist-:	Det: 100 1					
				_								,					- Jacob						7.4									
																						_										