

# Training Configuration – Modelli a Confronto

Modello	Grad. Accum.	Batch Effettivo	Multi-GPU
Gemma T9	1	2	✓ (torch.distributed)
Gemma T8	8	16 (2x8)	✓ (torch.distributed)
LLaMA T9	1	2	✓ (torch.distributed)
LLaMA T8	8	16 (2x8)	✓ (torch.distributed)

Tutti: 50k steps, eval/save ogni 250, max length 2,048 token

Epochs: calcolate automaticamente

### Optimizer & Learning Rate

Parametro	Valore
Optimizer	AdamW
Learning Rate	5e-5 ( $5 \times 10^{-5}$ )
Scheduler	Cosine (warmup 1k steps, ratio 0.02)
Weight Decay	0.01
Adam $\beta$	(0.9, 0.999)
Max Grad Norm	1.0

### LoRA Configuration

Parametro	Valore
Rank / Alpha	64 / 128
Dropout	0.05
Target Modules	q/k/v/o_proj, gate, up/down_proj
PEFT Type	LORA
Alpha/Rank Ratio	2.0

### Quantizzazione (Tutti i Modelli)

Parametro	Valore
4bit	✓
Quant Type	nf4
Compute Dtype	float16
Double Quant	✓
Storage	uint8

### Dataset Configuration

Parametro	Valore
Train/Val File	90k / 10k esempi
Formato Input	SVG XML + Caption in linguaggio naturale
Max SVG / Caption Length	1,800 / 200 caratteri
Preprocessing	Tokenizzazione dinamica, padding e truncation attivi

### DeepSpeed Configuration

Stage	Risparmio Memoria	Impatto Velocità
2	~50%	~15% più lento
3	~80–90%	~40% più lento

# Confronto modelli baseline

