**Algorithm 1** Pseudo code for Grounding-Bench Evaluation

```
    # Inputs:  Grounded response (T, b); GT response and
      boxes (T̂, b̂).
    # Outputs:  Chat scores (S); Grounded response scores
      (P, R, F).
    # Functions:  GPT_chat_eval(), match_gt_box(),
      construct_gd_response(), GPT_gd_eval(), calc_scores().
1  def GroundingBenchEval(T, b, T̂, b̂):
2      S = GPT_chat_eval(T, T̂) # Request GPT-4 to score T
3      N=len(b̂),  N_=len(b)
4      b_ = []
5      for b in b:
6          gt_box, iou = match_gt_box(b, b̂)
7          if iou ≥ 0.5:
8              b_=b_+gt_box
9      T_ = construct_gd_response(T, b_)# Insert matched
         gt boxes into the response to form responses of
         Context type 3 in Table 2
10     TP_,TP = GPT_gd_eval(T_)         # GPT help evaluate
         semantic matching
11     P, R, F = calc_scores(TP_,TP,N)  # Calculate
         grounded response scores
12     output = S, P, R, F
13     return output
```